

Local Linear Regression (LLR) for Pose Invariant Face Recognition

Xiujuan Chai^{1,2}, Shiguang Shan², Xilin Chen², Wen Gao^{1,2}

¹*School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China*

²*ICT-ISVISION FRJDL, Institute of Computing Technology, CAS, Beijing, 100080, China*
{xjchai, sgshan, xlchen, wgao}@jdl.ac.cn

Abstract

The variation of facial appearance due to the viewpoint (pose) degrades face recognition systems considerably, which is well known as one of the bottlenecks in face recognition. One of the possible solutions is generating virtual frontal view from any given non-frontal view to obtain a virtual gallery / probe face. By formulating this kind of solutions as a prediction problem, this paper proposes a simple but efficient novel Local Linear Regression (LLR) method, which can generate the virtual frontal view from a given non-frontal face image. The proposed LLR inspires from the observation that the corresponding local facial regions of the frontal and non-frontal view pair satisfy linear assumption much better than the whole face region. This can be explained easily by the fact that a 3D face shape is composed of many local planar surfaces, which satisfy naturally linear model under imaging projection. In LLR, we simply partition the whole non-frontal face image into multiple local patches and apply linear regression to each patch for the prediction of its virtual frontal patch. Comparing with other methods, the experimental results on CMU PIE database show distinct advantage of the proposed method. .

1. Introduction

Face recognition has been extensively studied for more than three decades. So far, the state-of-the-art recognition technology can achieve very high accuracy under restricted environment, such as frontal faces with indoor lighting conditions [1]. However, for those uncontrolled cases (e.g. outdoor with uncooperative subjects), the face recognition task still far from the requirements, since most of current face recognition systems are pretty sensitive to the pose, lighting, and other variations.

The difference between two poses will induce the large variation of the appearance even for the same person. The distinction is often more remarkable than that caused by

the difference of identity under the same pose. Therefore, the conventional appearance-based methods, such as Eigenface, will degrade dramatically when the input image is non-frontal while matching against frontal face images.

Many approaches have been proposed to recognize faces under various poses. Among them, the view-based methods are widely used [2-7]. For instance, view-based Eigenface had been proposed to extend the Eigenface to deal with the pose problem. But, view-based methods generally need multiple face images with different pose for each person in the database, which is often impractical for real world applications.

Gross et. al. proposed the Eigen Light-Fields (ELF) method to tackle the pose problem[8, 9]. This algorithm operates by estimating the Eigen light-fields of the subject's head from the input images. Matching between the probe and gallery is then performed by means of the eigen light-fields. The advantage is that only training set needs multiple samples for each person, which are easy to collect. But the precise computation of the plenoptic function is difficult, and in [8, 9], the authors substituted the plenoptic function approximately by concatenating the normalized image vectors under different poses.

To achieve the pose-invariant face recognition, another popular solution is to generate virtual views to normalize pose or expand the gallery. For the image appearance variations caused by pose variation have close relation to the 3D face structure, some researchers explored 3D model-based method to recover the intrinsic shape or albedo parameters [10-14] for classification. Having the specific 3D model, the novel face under any viewpoint can be easily obtained by rendering with graphics technique. The most representative method is the 3D morphable model method [11, 12]. This method has shown its good performance on FRVT 2002 [1]. However this is a high computation consumption procedure and it's too slow to be used in real-time tasks.

There also exist some learning-based view synthesis methods which can generate virtual views under multiple poses [15-18]. Beymer and Poggio et al proposed an

algorithm to synthesize novel views from single image by using the prior knowledge of the face images and applied them to face recognition. In their method, a face image was separated into shape vector and texture vector, and the Linear Object Classes (LOC) is applied to them respectively. Then the virtual “rotated” images are generated easily using a basis set of 2D prototypical views. The reconstructed virtual views are highly depends on the segmentation of the given face image and the correspondence between the images. However, building accurate pixel-wise correspondence between face images is an unsolved hard problem, which has prevented this method from further practical applications.

In this paper, we start from the basic idea of Linear Object Class, that is, we try to predict the frontal view of a given non-frontal face image using regression method based on the prototypes in a training set with corresponding image pairs of some specific poses. However, totally unlike Beymer’s method, we operate the regression algorithm directly on the image rather than the separated shape and texture. Therefore, the intractable accurate dense correspondence between face images is not required any more. What we need is just a coarse alignment based on the two irises. In the proposed Local Linear Regression (LLR) method, we partition the whole non-frontal face image into multiple local patches and apply linear regression to each patch for the prediction of its frontal patch. This is inspired from the fact that a 3D face shape is composed of many local planar surfaces, which satisfy naturally linear model under imaging projection. Compared with the LOC approach, LLR is simpler and easier for real world problem, because LLR does not require accurate shape and texture alignment, which are mandatory in LOC.

The remaining parts of the paper are organized as follows: Section 2 proposes the LLR-based virtual view generation approach. Section 3 presents the experimental results on the virtual view generation and pose-invariant face recognition. Section 4 is the conclusion.

2. Virtual view generation based on local linear regression

In this section, we first formulate the virtual view generation task as a general prediction framework. Then the global linear regression and local linear regression are presented respectively to predict virtual frontal view from a non-frontal face image.

2.1. Problem formulation and solution

Given a non-frontal facial image, our aim is to generate its virtual frontal view based on a training set. We formulate this problem mathematically as a regression task. Formally, it is described as follows:

Let $\{(\mathbf{X}^{P_0}, \mathbf{X}^{P_k})\}$ be the training set, which is composed of the corresponding samples under frontal pose P_0 and some specific non-frontal pose P_k . Here, $\mathbf{X}^{P_0} = (\mathbf{x}_1^{P_0}, \mathbf{x}_2^{P_0}, \dots, \mathbf{x}_N^{P_0})$ denotes the frontal face set composed of N subjects, and $\mathbf{X}^{P_k} = (\mathbf{x}_1^{P_k}, \mathbf{x}_2^{P_k}, \dots, \mathbf{x}_N^{P_k})$ is the corresponding non-frontal face set under pose P_k . Note that $\mathbf{x}_i^{P_k}$ is the counterpart image of $\mathbf{x}_i^{P_0}$ from the same person but with different pose. For the facial shape is similar in holistic, and the albedo of a face can be assumed to be a constant [19], there should be a mapping

$$f: \mathbf{x}^{P_k} \rightarrow \mathbf{x}^{P_0} \quad (1)$$

which can transform a non-frontal face \mathbf{x}^{P_k} into its frontal counterpart \mathbf{x}^{P_0} . Though, imaginably, the mapping could be very complicated (non-linear), in this paper we only consider the solution under linear assumption. In this case, the mapping can be rewritten as:

$$\mathbf{x}^{P_0} = \mathbf{A} \mathbf{x}^{P_k}, \quad (2)$$

where \mathbf{A} is a linear operator. Let n denote the number of pixels of an image, and a training set with N corresponding images as above mentioned. If $n > N$, the linear mapping \mathbf{A} can be estimated by the following linear regression procedure (least square solution):

$$\mathbf{A} = \mathbf{X}^{P_0} (\mathbf{X}^{P_k})^\perp, \quad (3)$$

where

$$(\mathbf{X}^{P_k})^\perp = \left((\mathbf{X}^{P_k}) (\mathbf{X}^{P_k})^\top \right)^{-1} (\mathbf{X}^{P_k})^\top \quad (4)$$

is the pseudo inverse of \mathbf{X}^{P_k} .

Once the linear mapping function \mathbf{A} is already estimated based on the training set, when given any image \mathbf{x}^{P_k} with pose P_k , its corresponding virtual frontal image \mathbf{x}^{P_0} can be computed according to the same linear transformation:

$$\mathbf{x}^{P_0} = \mathbf{A} \mathbf{x}^{P_k} = \mathbf{X}^{P_0} (\mathbf{X}^{P_k})^\perp \mathbf{x}^{P_k}. \quad (5)$$

Considering the virtual view generation in a step further, we rewrite the equation (5) as:

$$\mathbf{x}^{P_0} = \mathbf{X}^{P_0} \boldsymbol{\alpha}, \quad (6)$$

where

$$\boldsymbol{\alpha} = (\mathbf{X}^{P_k})^\perp \mathbf{x}^{P_k}. \quad (7)$$

Therefore the virtual view generation can be decomposed into two steps: one is the solving of reconstruction coefficients in the P_k pose image space by equation (7); the other is the final virtual frontal view prediction using equation (6). Actually, the above solution of the reconstruction coefficients is the result of an optimization procedure aiming at seeking a coefficients vector, which can best represent the input image in the P_k pose image space. This is achieved by minimizing the following residue function:

$$\varepsilon(\boldsymbol{\alpha}) = \|\mathbf{x}^{P_k} - \mathbf{x}_{\text{Rec}}^{P_k}\|^2, \quad (8)$$

where

$$\mathbf{x}_{\text{Rec}}^{P_k} = \mathbf{X}^{P_k} \boldsymbol{\alpha} = \sum_{j=1}^N \mathbf{x}_j^{P_k} \alpha_j, \quad (9)$$

is the projection of \mathbf{x}^{P_k} in the P_k pose image space. Hereinafter, it is called reconstructed image.

In addition, from the above analysis, one can understand the linear regression more clearly as follows: the virtual frontal view of an input non-frontal face image \mathbf{x}^{P_k} with pose P_k is generated through a linear combination by using the same coefficients reconstructing the \mathbf{x}^{P_k} in the P_k pose image space. Coincidentally, this idea is quite consistent with the concept of linear object class [18], but we have formulated the problem quite differently from the point of view of regression.

2.2. Global linear regression

Based on the above analysis, we can easily derive the virtual frontal view from the single non-frontal facial image by using equation (5). Note that, when this procedure is implemented, one should carefully align the face images. As is well accepted in face recognition area, one can simply just align the faces according to the eye centers. And then the normalized face images in a whole are used to feed into the above prediction. We call this implementation as Global Linear Regression (GLR).

However the face is not planar in whole, that is to say, the absolute linear mapping between two different views of a person does not exist. Therefore, both the reconstruction of the input image in pose image space and the prediction in the frontal image space are not as precise as expected. Please refer to the (b) columns of Fig 2 and Fig.3 in section 3 for some reconstruction and prediction examples based on GLR, from which one can see obvious ghost and blur effect.

Considering that some facial patch is more like a planar surface, a natural improvement of GLR is applying linear regression locally.

2.3. Local linear regression

The intrinsic non-planar structure of the face makes the linear assumption fails when the facial pose changes. It also degrades the results for the virtual view generation based on GLR. Considering that a 3D face surface is composed of many local planar regions, for each small patch, the linear mapping will be maintained better both in the single pose image space and across different poses than the global case. So we further propose a method to synthesize virtual views by Local Linear Regression (LLR), in which linear regression is conducted in patch-wise mode. Concretely, face images are partitioned into uniformed blocks, and then each block are predicted using linear regression, as illustrated in Fig.1. This procedure is formally formulated as follows:

Firstly, given the training set, we need partition each face image into M blocks. Due to the pose variation, different partition modes are performed to the images according to their pose categories. In our method, the frontal faces are partitioned into regular grids, while the partitioning of images with P_k pose is completed by coarsely seeking for the counterpart of the frontal patches by the aid of an average 3D face model. This ensures the corresponding local patches in frontal and pose image possess the same semantics, as can be seen from Fig.1.

Then, given an input image \mathbf{x}^{P_k} whose pose is P_k , we partition it into M small patches $\mathbf{x}^{P_k} = (\mathbf{x}^{(1,P_k)} \quad \mathbf{x}^{(2,P_k)} \quad \dots \quad \mathbf{x}^{(M,P_k)})$ as is done for the P_k pose training images. Predicting the corresponding i -th frontal patch $\mathbf{x}^{(i,P_0)}$ for the i -th non-frontal patch $\mathbf{x}^{(i,P_k)}$ follows two steps:

- [1] First, estimate the reconstruction coefficients for the i -th small input patch in the specific patch space by:

$$\boldsymbol{\alpha}^i = (\mathbf{X}^{(i,P_k)})^\dagger \mathbf{x}^{(i,P_k)}, \quad (10)$$

where $\mathbf{X}^{(i,P_k)} = (\mathbf{x}_1^{(i,P_k)} \quad \mathbf{x}_2^{(i,P_k)} \quad \dots \quad \mathbf{x}_N^{(i,P_k)})$ is the i -th patches with P_k pose from the training set.

- [2] Then, the virtual frontal patch can be gotten by:

$$\mathbf{x}^{(i,P_0)} = \mathbf{X}^{(i,P_0)} \boldsymbol{\alpha}^i, \quad (11)$$

where $\mathbf{X}^{(i,P_0)} = (\mathbf{x}_1^{(i,P_0)} \quad \mathbf{x}_2^{(i,P_0)} \quad \dots \quad \mathbf{x}_N^{(i,P_0)})$ is the i -th patch from the frontal images in the training set.

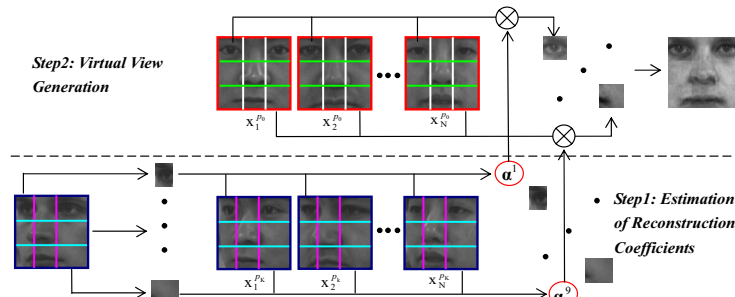


Fig.1. The flow chart of our proposed virtual view generation method based on LLR.

After performing such prediction for each patch in the \mathbf{x}^{P_k} , we combine all the small virtual frontal patches into a whole vector, that is the target virtual frontal view $\mathbf{x}^{P_0} = (\mathbf{x}^{(1,P_0)} \ \mathbf{x}^{(2,P_0)} \ \dots \ \mathbf{x}^{(M,P_0)})$. The resulting normalized frontal view is then used for recognition.

3. Experimental results

In this section, we conduct experiments on the 4 pose subsets of CMU PIE database, which includes the pose 29, 05 (turning left and right at 22.5 degree) and 11, 37 (turning left and right at 45 degree)[20]. In our experiments, leave-one-out strategy is used for generating the virtual frontal views. The final face recognition is performed on the totally 68 subjects, with the images of pose 27 forming the gallery matching with the virtual views.

3.1. The experimental results on the virtual view generation

In this section, some virtual view generation results are presented. The virtual frontal images are generated for all the images in 4 non-frontal pose sets. For the virtual view generation can be decomposed into input image reconstruction (Eq.9) and the virtual view prediction (Eq. 6 for GLR and Eq.11 for LLR), the results for the two stages are shown in Fig.2 and Fig.3 respectively.

In Fig.2 the first column is the input non-frontal images from 4 different poses, which are the pose sets 29, 05, 11 and 37 from top to down. The column (b) shows the reconstruction results of GLR and the rest columns (c) through (g) give the results of LLR with different patch size labeled at the top of the figure. Note that this size is that for the frontal face patches and the size for the non-frontal patch is accordingly changed. The corresponding virtual view prediction results are given in Fig.3, where the first column (a) is the input images. Column (b) shows the virtual view prediction results of GLR, and the columns (c)-(g) illustrate the results of LLR with different patch size. The last column (h) gives the ground truth.

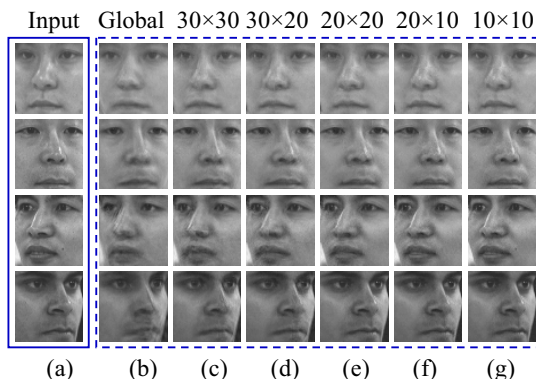


Fig.2. Some input non-frontal images reconstruction results. The first column is the input faces under 4 different poses. The rights are the reconstructed faces of GLR and LLR respectively.

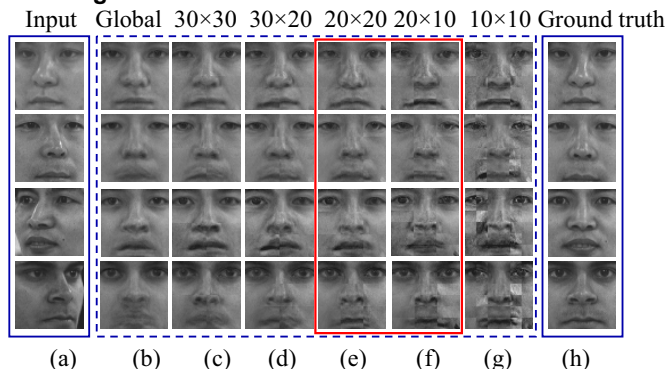


Fig.3. Examples of the virtual frontal view prediction results. First column is the input non-frontal facial images. Column (b) is the virtual frontal views by GLR and columns (c)-(g) are the results with LLR on 5 different scales of patch size. The last column is the real frontal faces.

From these results, it is obvious that, both the input face reconstruction and the virtual frontal view prediction of GLR are somewhat blurred with some ghost effect. While, LLR makes the linear combination across poses more reasonable since the better local linearity is maintained than the global case. But if the piece size is too small, the coarse correspondence across different poses for each patch will be meaningless which results in the bad prediction results as shown in Fig.3 column (g). As indicated by Fig.3, the best virtual frontal views are achieved for the appropriate piece size 20×20 or 20×10 , which are shown on the columns (e) and (f). In these cases, the local information of face, especially the eye region, is well recovered compared with the global reconstruction and other partitions by a too large or small size.

3.2. The experimental results on pose invariant face recognition

In this section, pose invariant face recognition experiments are carried out on the virtual frontal views to evaluate the proposed algorithm.

Based on the pose normalization results, we found that the upper faces are reconstructed more vivid than the lower faces. This is because when we partition face, the upper will get better correspondence for having more similar and planar 3D face structure. Also, the preprocessing to normalize image by fixing the eyes location makes the upper alignment more accurate. On the contrary, the nose region varies to a large extent with the complex geometric structure for each person. The rotation transformation for the lower region is away from linearity, which leads to the reconstruction bias for the lower face.

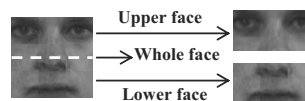


Fig.4. The face region dividing for recognition.

Through above analysis, recognition is designed only for the upper face as Fig. 4 illustrated. The experiential borderline is used to get valuable region. That is to say, only the upper is selected for both probe image and gallery images. Then the correlations are calculated between the upper faces of the probe vector and each gallery vector. The subject with the maximal correlation is the recognition result. Fig.5 gives the recognition results for the 4 pose sets of PIE database with virtual frontal views generated by both GLR and LLR. The recognition result for the uppers of the original pose images is regarded as the baseline. From the results, one can find that better recognition results are achieved on the patch size being 20×20 or 20×10 . This is consistent with the visualized virtual view generation results. Also, the comparison between the recognition using different face region is presented in Fig.6, in which, the upper face, lower face and the whole face are considered respectively. The baseline is the recognition result using the corresponding region of the original pose images.

The Eigen Light Fields (ELF) algorithm is well known for recognizing faces across pose and achieving good performance [9]. We compare our results to those of ELF with the 3-point normalization on the same pose subsets of PIE database. In ELF, half of the subjects are randomly selected for training and the recognition is performed on the rest 34 subjects. Taking the pose set 27 as the gallery, our method achieves better result than ELF as table 1 illustrated.

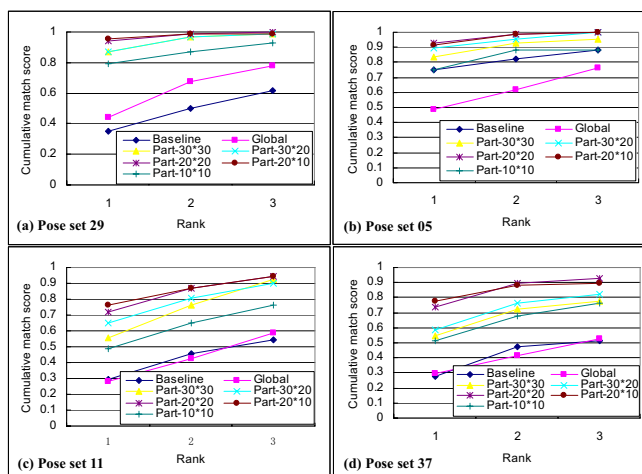


Fig.5. Recognition results on the original image and the pose normalized images on 4 pose sets of PIE database with GLR and LLR.

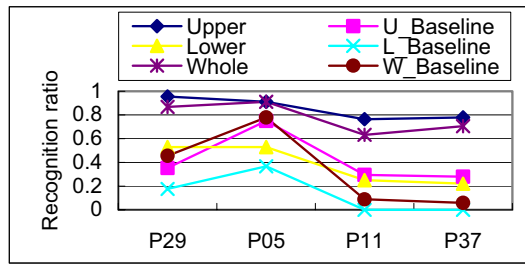


Fig.6. The comparison of the recognition results using different face region.

Table 1. The performance comparison between our method and the ELF.

Method \ Pose Set	Our Method (Patch size 20×10)	ELF method [9] (3-P Normalization)
Pose set 29	95.6%	86%
Pose set 05	91.2%	88%
Pose set 11	76.5%	76%
Pose set 37	77.9%	74%

4. Conclusion and discussions

By formulating the virtual frontal view generation as a prediction problem, we propose the LLR-based approach to generate virtual frontal view for pose invariant face recognition. The experimental results show that the local linearity of semantically corresponding patches between frontal and non-frontal face images can greatly facilitate the prediction. The effectiveness of the proposed approach has been validated by the recognition experiments on the PIE faces with up to 45 degree rotation in depth.

In addition, compared with other methods, such as Eigen light-field, 3D morphable model, the LLR-based approach is more robust to alignment except the two irises.

In this paper, we model the prediction under the piecewise linear assumption. However, it is in nature nonlinear. Therefore, we will extend LLR to non-linear mapping for better prediction result in the future.

Acknowledgement

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, "100 Talents Program" of CAS, the Program for New Century Excellent Talents in University (NCET-04-0320) and ISVISION Technologies Co., Ltd.

References

- [1] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi and M. Bone, "Face Recognition Vendor Test 2002: Evaluation Report", FRVT 2002, March 2003.
- [2] M. Turk and A. Pentland, "Eigenfaces for Recognition", *Journal of Cog. Neuroscience*, 1991, 3(1): 71-96.
- [3] H. Murase and S. Nayar, "Learning and Recognition of 3D Objects from Appearance", *Proc. Qualitative Vision Workshop*, New York, June, 1993. pp.39-50.
- [4] A. Pentland, B. Moghaddam and T. Starner, "View-based and Modular Eigenspace for Face Recognition", *CVPR*, 1994, pp. 84-91.
- [5] H. Murase and S.K. Nayar, "Visual Learning and Recognition of 3-D Objects form Appearance", *IJCV*, 1995, 14(1):5-24.
- [6] S. McKenna, S. Gong and J. Collins, "Face Tracking and Pose Representation", *BMVC*, Edinburgh, Scotland, 1996, pp. 755-764.
- [7] Z. Zhou, J. HuangFu, H. Zhang and Z. Chen, "Neural Network Ensemble Based View Invariant Face Recognition", *Journal of Computer Study and Development*, 2001, 38(9):1061-1065.
- [8] R. Gross, I. Matthews, and S. Baker, "Eigen Light-Fields and Face Recognition Across Pose", *FGR*, Washington DC, 2002, pp.3-9.
- [9] R. Gross, I. Matthews, and S. Baker, "Appearance-Based Face Recognition and Light-Fields", *IEEE Trans. On PAMI*, 2004, 26:449-465.
- [10] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang and W. Gao, "Efficient 3D Reconstruction for Face Reconstruction", *PR.*, 2005, 38(6): 787-798.
- [11] V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D Faces", *In Proceedings, SIGGRAPH'99*, 1999, pp. 187-194.
- [12] V. Blanz and T. Vetter, "Face Recognition based on Fitting a 3D Morphable Model", *IEEE Transactions on PAMI*, 2003, 25: 1063-1074.
- [13] W. Zhao and R. Chellappa, "SFS Based View Synthesis for Robust Face Recognition", *FGR*, Grenoble, 2000, pp. 285-292.
- [14] A.S. Georghiadis, P.N. Belhumeur and D.J. Keigman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Poses", *IEEE Transactions on PAMI*, 2001, 23:643-660.
- [15] D. Beymer and T. Poggio, "Face Recognition from One Example View," *ICCV*, 1995, pp. 500-507.
- [16] D. Beymer, "Face Recognition under Varying Pose", *Technical Report A.I.Memo*, No.1461, Artificial Intelligence Laboratory, MIT. 1993.
- [17] T. Vetter, "Synthesis of Novel Views from a Single Face Image", *IJCV*, 1998, 28(2): 103-116.
- [18] T. Vetter and T. Poggio, "Linear Object Classes and Image Synthesis from a Single Example Image," *IEEE Trans. On PAMI*, 1997, 19(7): 733-742
- [19] Z. Wen, Z. Liu and T.S. Huang, "Face Relighting with Radiance Environment Maps," *CVPR*, 2003, pp. 158-165.
- [20] T. Sim, S. Baker and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database", *FGR'02*, Washington, DC, 2002, pp. 46-51.