Converting Thermal Infrared Face Images into Normal Gray-Level Images

Mingsong Dou[1], Chao Zhang[1], Pengwei Hao[1,2], and Jun Li[2]

[1] State Key Laboratory of Machine Perception, Peking University, Beijing, 100871, China
[2] Department of Computer Science, Queen Mary, University of London, E1 4NS, UK
c.zhang@pku.edu.cn

Abstract. In this paper, we address the problem of producing visible spectrum facial images as we normally see by using thermal infrared images. We apply Canonical Correlation Analysis (CCA) to extract the features, converting a many-to-many mapping between infrared and visible images into a one-to-one mapping approximately. Then we learn the relationship between two feature spaces in which the visible features are inferred from the corresponding infrared features using Locally-Linear Regression (LLR) or, what is called, Sophisticated LLE, and a Locally Linear Embedding (LLE) method is used to recover a visible image from the inferred features, recovering some information lost in the infrared image. Experiments demonstrate that our method maintains the global facial structure and infers many local facial details from the thermal infrared images.

1 Introduction

Human facial images have been widely used in the biometrics, law enforcement, surveillance and so on [1], but only the visible spectrum images of human faces were used in most cases. Recently the literature begins to emerge for face recognition (FR) based on infrared images or fusion of infrared images and visible spectrum images [2-4], and some sound results have been published.

Other than FR based on infrared images this paper focuses on the transformation from thermal IR images to visible spectrum images (see Fig.2 for examples of both modal images), i.e. we try to render a visible spectrum image from a given thermal infrared image.

Thermal infrared imaging sensors measure temperature of shot objects and are invariant to illuminance. There are many surveillance applications in which the light conditions are so poor that we can only acquire thermal infrared images. As we know, we see objects because of the reflectance of light, i.e. formation of visible-spectrum images needs light sources. For thermal infrared images, it is very optimistic. All objects with temperature above the absolute zero emit electromagnetic wave, and the human body temperature is in the range of emitting infrared electromagnetic wave. So even it is completely dark, we can still obtain thermal infrared images with thermal infrared imaging cameras.

Though the formations of visible spectrum and infrared facial images are of different mechanisms, the images do share some commons if they come from the same face, e.g.

we can recognize some facial features from both modals of images. There indeed exists some correlation relationship between them which can be learned from training sets.

The problem to normally view thermal infrared images is actually very challenging. First of all, the correlations between the visible image and corresponding infrared one are not strong. As mentioned above, the imaging models are of different mechanisms. Infrared images are invariant under the changes of the lighting conditions, so many visible spectrum images taken under different lighting conditions correspond to one infrared image. Therefore, the solution to our problem is not unique. To the contrary, thermal infrared images are not constant, either. Thermal infrared images are subject to the surface temperature of the shot objects. For example, the infrared images taken respectively from a person when he just came from the cold outside and the same person when he just did lots of sports are quite different. The analysis above shows that it is a many-to-many mapping between visual facial images and thermal infrared images of the same person, which is the biggest barrier for solving the problem. Another problem is that the resolution of visible spectrum facial images is generally much higher than that of thermal infrared images. Thus visible images have more information, and some information of visible spectrum images definitely can not be recovered from thermal infrared images through the correlation relationship.

In this paper, we have developed a method to solve the problems. We use Canonical Correlation Analysis (CCA) to the extract features, converting a many-to-many mapping between infrared and visible image into a one-to-one mapping approximately. Then we learn the relationship between the feature spaces, in which the visible features are inferred from the corresponding infrared features using Locally-Linear Regression (LLR) or, what is called, Sophisticated LLE, and a Locally Linear Embedding (LLE) method is applied to recover a visible image from the inferred features, recovering some information lost in the infrared image.

2   Related Works

As presented above, this paper addresses the problem of conversion between different modal images, which shares lots of commons with the super-resolution problem [5-7], which is to render one high resolution (HR) image from one or several low resolution (LR) images. For example, the data we try to recover for two problems both have some information lost in the given observation data.

Baker et al. [5] developed a super-resolution method called face hallucination to recover the lost information. They first matched the input LR image to those in the training set, found the most similar LR image, and then take the first derivation information of the corresponding HR image in the training set as the information of the desired HR image. We adopt this idea of finding information from the training set for the recovery data.

Chang et al. [6] introduced LLE [10] to super-resolution. Their method is based on the assumption that the patches in the low- and high- resolution images form the manifolds with the same local geometry in two distinct spaces, i.e. we can reconstruct a HR patch from the neighboring HR patches with the same coefficients as that for re-constructing the corresponding LR patch from the neighboring LR patches. Actually this method is a special case of Locally-Weighted Regression (LWR) [8] when the

weights for all neighbors are equal and the regression function is linear, as we show in Section 4. We develop a Sophisticated LLE method which is an extension of LLE.

Freeman et al. [7] took images as a Markov Random Field (MRF) with the nodes corresponding to image patches, i.e. the information from the surrounding patches is used to constrain the solution, while LLE does not. MRF improves the results when the images are not well aligned, but in our paper we assume all the images are well-registered and we do not use the time-consuming MRF method.

Our work is also related to some researches on statistical learning. Melzer et al. [11] used Canonical Correlation Analysis (CCA) to infer the pose of the object from the gray-level images, and Reiter et al.'s method [12] learns the depth information from RGB images also using CCA. CCA aims to find two sets of projection directions for two training sets, with the property that the correlation between the projected data is maximized. We use CCA for feature extraction.

Shakhnarovich et al. [9] used Locally-Weighted Regression for pose estimation. To accelerate searching for the nearest neighbors (NN), they adopted and extended the Locality-Sensitive Hashing (LSH) technique. The problem we address here is much different from theirs. A visible image is not an underlying scene to generate an infrared image, while in their problem the pose is the underlying parameter for the corresponding image. So we use CCA to extract the most correlated features, at the same time the dimensionality of data is reduced dramatically, making nearest neighbors searching easier. In our experiments we use exhaustive search for NN instead of LSH.

3   Feature Extraction Using CCA

As mentioned above, the correspondence between the visible and the infrared images is a many-to-many mapping, and then to learn a simple linear relationship between the two image spaces is not possible. Instead, extracting features and learning the relationship between the feature spaces can be a solution. We wish to extract features from the original image with the properties as follows: (1) The relationship between two feature spaces is stable, i.e. there exists a one-to-one mapping between them, and it is easy to be learned from the training set and performs well when generalized to the test set; (2) The features in two distinct feature spaces should contain enough information to approximately recover the images. Unfortunately for our problem the two properties conflict with each other.

Principal Component Analysis (PCA), which is known as the EigenFace method [13] in face recognition, is a popular method to extract features. For our problem it well satisfies the second condition above, but two sets of principal components, extracted from a visible image and the corresponding infrared image, have weak correlations.

Canonical Correlation Analysis (CCA) finds pairs of directions that yield the maximum correlations between two data sets or two random vectors, i.e. the correlations between the projections (features) of the original data projected onto these directions are maximized. CCA has our desired traits as given in the above property (1). But unlike PCA, several CCA projections are not sufficient to recover the original data, for the found directions may not be able to cover the principal variance of the data set. However, we find that regularized CCA is a satisfying trade-off between the two desired properties.

3.1 Definition of CCA

Given two zero-mean random variables $\mathbf{x}$, a $p \times 1$ vector, and $\mathbf{y}$, a $q \times 1$ vector, CCA finds the *1*st pair of directions $\mathbf{w}_1$ and $\mathbf{v}_1$ that maximize the correlation between the projections $x = \mathbf{w}_1^T\mathbf{x}$ and $y = \mathbf{v}_1^T\mathbf{y}$,

$$\max \rho(\mathbf{w}_1^T\mathbf{x}, \mathbf{v}_1^T\mathbf{y}) \,,$$

$$s.t. \;\; Var(\mathbf{w}_1^T\mathbf{x}) = 1 \text{ and } Var(\mathbf{v}_1^T\mathbf{y}) = 1 \,, \tag{1}$$

where ρ is the correlation coefficient, the variables x and y are called the first canonical variates, and the vectors $\mathbf{w}_1$ and $\mathbf{v}_1$ are the first correlation direction vector. CCA finds *k*th pair of directions $\mathbf{w}_k$ and $\mathbf{v}_k$ satisfying: (1) $\mathbf{w}_k^T\mathbf{x}$ and $\mathbf{v}_k^T\mathbf{y}$ are uncorrelated to the former *k-1* canonical variates; (2) the correlation between $\mathbf{w}_k^T\mathbf{x}$ and $\mathbf{v}_k^T\mathbf{y}$ is maximized subject to the constraints $Var(\mathbf{w}_k^T\mathbf{x}) = 1$ and $Var(\mathbf{v}_k^T\mathbf{y}) = 1$. Then $\mathbf{w}_k^T\mathbf{x}$ and $\mathbf{v}_k^T\mathbf{y}$ are called the *k*th canonical variates, and $\mathbf{w}_k$ and $\mathbf{v}_k$ are the *k*th correlation direction vector, $k \leq \min(p, q)$.

The solution for the correlation directions and correlation coefficients is equivalent to the solution of the generalized eigenvalue problem below,

$$(\textstyle\sum_{xy}\sum_{yy}^{-1}\sum_{xy}^T - \rho^2\sum_{xx})\mathbf{w} = 0 \,, \tag{2}$$

$$(\textstyle\sum_{xy}^T\sum_{xx}^{-1}\sum_{xy} - \rho^2\sum_{yy})\mathbf{v} = 0 \,, \tag{3}$$

where $\sum_{xx}$ and $\sum_{yy}$ are the self-correlation matrices, $\sum_{xy}$ and $\sum_{yx}$ are the co-correlation matrices. There are robust methods to solve this problem, interested readers please refer to [15], where an SVD-based method is introduced.

Unlike PCA, which aims to minimize the reconstruction error, CCA puts the first place the correlation of the two data sets. There is no assurance that the directions found by CCA cover the main variance of the data set, so generally speaking a few projections (canonical variates) are not sufficient to recover the original data well.

Beside the recovery problem, we also have to deal with the overfitting problem. CCA is sensitive to noise. Even if there is small amount of noise in the data, CCA might give a good result to maximize the correlations between the extracted features, but the features more likely represent the noise rather than the data.

As mentioned in [11], it is a sound method to add a multiple of the identity matrix $\lambda\mathbf{I}$ to the co-variance matrix $\sum_{xx}$ and $\sum_{yy}$ to overcome the overfitting problem, and this method is called regularized CCA. We find that it also has effect on the reconstruction accuracy, as depicted in Fig.1. Then regularized CCA is a trade-off between the two desired properties mentioned above.

3.2 Feature Extraction and Image Recovery from Features

We extract the local features rather than the holistic features for the holistic features seem to fail to capture the local facial traits. There is a training set consisting of pairs of visible and infrared images at our disposal. We partition all the images into overlapping patches, then at every patch position we have a set of patch pairs for CCA learning, and CCA finds pairs of directions $\mathbf{W}^{(i)} = [\mathbf{w}_1,\mathbf{w}_2,...,\mathbf{w}_k]$ and $\mathbf{V}^{(i)} = [\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_k]$ for visible and infrared patches respectively, where the superscript (i) denotes the patch index

(or the patch position in the image). Every column of **W** or **V** is a unitary direction vector, but it is not orthogonal between different columns. Take a visible patch **p** (represented as a column vector by raster scan) at position i for example, we can extract the CCA feature of the patch **p**, using

$$\mathbf{f} = \mathbf{W}^{(i)\mathrm{T}}\mathbf{p} \, , \tag{4}$$

where **f** is the feature vector of the patch.

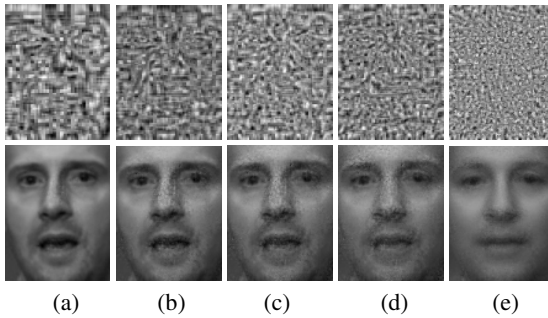(a)   (b)   (c)   (d)   (e)

Fig. 1. The first row is the first CCA directions with different λ (we rearrange the direction vector as an image, and there are outlined faces in the former several images), and the second row is the corresponding reconstruction results. CCA is patch-based as introduced in Section 3.2; we reconstruct the image with 20 CCA variates using Eq(6). If the largest singular value of variance matrix is c, we set (a) $\lambda = c/20$; (b) $\lambda = c/100$; (c) $\lambda = c/200$; (d) $\lambda = c/500$; (e) $\lambda = c/5000$. It is obvious that when λ is small, the CCA direction tends to be noisy, and the reconstructed face tends to the mean face.

It is somewhat tricky to reconstruct the original patch **p** through feature vector **f**. Since **W** is not orthogonal, we cannot reconstruct the patch by **p** = **Wf** as we do in PCA. However we can solve the least squares problem below to obtain the original patch,

$$\mathbf{p} = \mathrm{arg_p} \ \mathbf{min} \ \|\mathbf{W}^{\mathrm{T}}\mathbf{p} - \mathbf{f}\|_2^2 \, , \tag{5}$$

or to add an energy constraint,

$$\mathbf{p} = \mathrm{arg_p} \ \mathbf{min} \ \|\mathbf{W}^{\mathrm{T}}\mathbf{p} - \mathbf{f}\|_2^2 + \|\mathbf{p}\|_2^2 \, . \tag{6}$$

The least squares problem above can be efficiently solved with the scaled conjugate gradient method.

The above reconstruction method is feasible only in the situation when the feature vector **f** contains enough information of the original patch. When fewer canonical variates (features) are extracted, we can recover the original path using LLE method [10]. As the method in [6], we assume that the manifold of the feature space and that of the patch space have the same local geometry; then the original patch and its features have the same reconstruction coefficients. If $\mathbf{p}_1, \mathbf{p}_2,..., \mathbf{p}_k$ are the patches whose features $\mathbf{f}_1, \mathbf{f}_2,..., \mathbf{f}_k$ are **f**'s k nearest neighbors, and **f** can be reconstructed from neighbors with **f** = **Fw**, where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2,..., \mathbf{f}_k]$, $\mathbf{w} = [w_1, w_2,..., w_k]^{\mathrm{T}}$, we can reconstruct the original patch by

$$\mathbf{p} = \mathbf{P}\mathbf{w} , \qquad (7)$$

where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2,\ldots, \mathbf{p}_k]$. The reconstruction results using Eq(6 & 7) are show in Fig. 3(**a**). When only a few canonical variates at hand, the method of Eq(7) performs better than Eq(6); while there are more canonical variates, two methods give almost the same satisfying results.

4   Facial Image Conversion Using CCA Features

From the training database we can obtain the CCA projection directions at every patch position, and for all the patches from all the training images we extract features by projection onto the proper directions, then at each patch position i we get a visible training set $\mathbf{O}_v^{\ i} = \{\mathbf{f}_{v,j}^{\ i}\}$ and an infrared one $\mathbf{O}_{ir}^{\ i} = \{\mathbf{f}_{ir,j}^{\ i}\}$.

Given one new infrared image, we partition it into small patches, and obtain the feature vector $\mathbf{f}_{ir}$ of every patch. If we can infer the corresponding visible feature vector $\mathbf{f}_v$, the visible patch can be obtained using Eq(7) and then the patches will be combined into an visible facial image. In this section we will focus on the prediction of the visible feature vector from the infrared one. Note that the inferences for patches at different positions are based on different training feature sets.

4.1   Reconstruction Through Locally-Linear Regression

Locally-Weighted Regression [8][9] is a method to fit a function of the independent variables locally based on a training set, and it suits our problem well. To simplify the methods, we set the weights of nearest neighbors (NN) equal, and use a linear model to fit the function, then LWR degenerates to Locally-Linear Regression (LLR).

For an input infrared feature vector $\mathbf{f}_{ir}$, we find K-NNs in training set $\mathbf{O}_{ir}$, which compose a matrix $\mathbf{F}_{ir} = [\mathbf{f}_{ir,1}, \mathbf{f}_{ir,2},\ldots, \mathbf{f}_{ir,K}]$, and their corresponding visible feature vectors compose a matrix $\mathbf{F}_v = [\mathbf{f}_{v,1}, \mathbf{f}_{v,2},\ldots, \mathbf{f}_{v,K}]$. Note that we omit the patch index for convenience. Then a linear regression obtains the relation matrix,

$$\mathbf{M} = \arg_{\mathbf{M}} \min \sum_k \| \mathbf{f}_{v,k} - \mathbf{M}\,\mathbf{f}_{ir,k}\|_2^{\ 2}$$
$$= \mathbf{F}_v.\,\mathbf{F}_{ir}^{\ +} , \qquad (8)$$

where $\mathbf{F}_{ir}^{\ +}$ is the pseudo-inverse of $\mathbf{F}_{ir}$. The corresponding visible feature vector $\mathbf{f}_v$ can be inferred from the input infrared feature $\mathbf{f}_{ir}$ by

$$\mathbf{f}_v = \mathbf{M}\,\mathbf{f}_{ir}$$
$$= \mathbf{F}_v\,\mathbf{F}_{ir}^{\ +}\,\mathbf{f}_{ir}. \qquad (9)$$

To find the nearest neighbors, the distance between two infrared feature vectors $\mathbf{f}^{\mathrm{T}}$ and $\mathbf{f}^{\mathrm{I}}$ need be defined. In this paper, we define the distance as

$$D = \sum_k \rho_k\, (f_k^{\ \mathrm{T}}, -f_k^{\ \mathrm{I}},) , \qquad (10)$$

where ρ_k is kth correlation coefficient; $f_k^{\ \mathrm{T}}$ and $f_k^{\ \mathrm{I}}$ denotes the kth element of the feature vector $\mathbf{f}^{\mathrm{T}}$ and $\mathbf{f}^{\mathrm{I}}$ respectively.

Actually the LLE method using in [6] is equivalent to LLR. The reconstruction coefficients $\mathbf{w}$ of infrared feature $\mathbf{f}_{ir}$ from K-NNs, $\mathbf{F}_{ir}$, can be obtained by solving the Least Squares Problem.

$$\mathbf{w} = \arg_{\mathbf{w}} \min \| \mathbf{F}_{ir} \, \mathbf{w} - \mathbf{f}_{ir} \|$$
$$= \mathbf{F}_{ir}^{+} \mathbf{f}_{ir} .$$

(11)

then the reconstructed corresponding visible feature vector $\mathbf{f}_v = \mathbf{F}_v \, \mathbf{F}_{ir}^{+} \mathbf{f}_{ir}$, which has the same form as Eq(9).

The difference between the two methods is the selection of the number K of NNs. In LLR, to make regression sensible we select a large K to ensure that $\mathbf{F}_{ir} \in \mathbf{R}^{m \times n}$ has more columns than rows (m<n), i.e. $\mathbf{F}_{ir}^{+} = \mathbf{F}_{ir}^{T}(\mathbf{F}_{ir} \, \mathbf{F}_{ir}^{T})^{-1}$. While in LLE method, a small K is selected, then m>n and $\mathbf{F}_{ir}^{+} = (\mathbf{F}_{ir}^{T} \mathbf{F}_{ir})^{-1} \, \mathbf{F}_{ir}^{T}$. The reconstruction results are shown in Fig.2(b)(e). The LLR method gives better results, but consumes more resources for it needs to find a larger number of NNs. In the next section we extend LLE to a Sophisticated LLE which achieves competitive results as LLR, while it uses approximately the same resources as LLE.

4.2 Reconstruction Via Sophisticated LLE

The reason of the poor performance of LLE method may be that the local geometry of the two manifolds of visible and infrared features is not the same. We use an experiment to demonstrate it. For every infrared feature vector $\mathbf{f}_{ir}^{i}$ in the training set we find its four NNs $\{\mathbf{f}_{ir}^{1}, \mathbf{f}_{ir}^{2}, \mathbf{f}_{ir}^{3}, \mathbf{f}_{ir}^{4}\}$ (the neighbors are organized in the decreasing order according to the distance to $\mathbf{f}_{ir}^{i}$, the same below.) whose convex hull (a tetrahedron) contains the infrared patch, but their visible counterpart, the visible feature vector $\mathbf{f}_v^{i}$ and its neighbors $\{\mathbf{f}_v^{1}, \mathbf{f}_v^{2}, \mathbf{f}_v^{3}, \mathbf{f}_v^{4}\}$ do not preserve the same geometric relations. Moreover, more than 90 percent of $\mathbf{f}_v^{i}$'s are out of the convex hulls of the corresponding neighbors.

It is a natural idea to learn the changes between two local geometries of two manifolds. Since the local geometry is represented by the reconstruction coefficients, we only need to learn the mapping $\mathbf{H}(\cdot)$ between the infrared and the visible reconstruction coefficients denoted as $\mathbf{x}$ and $\mathbf{y}$ respectively, and $\mathbf{y} = \mathbf{H}(\mathbf{x})$.

Since we have a training database at hand, we collect the pairs of reconstruction coefficient vectors $(\mathbf{x}_1, \mathbf{y}_1),\ldots,(\mathbf{x}_N, \mathbf{y}_N)$, which are used to reconstruct feature vectors of visible and infrared patches respectively. We can obtain the function $\mathbf{H}(\cdot)$ between them using the least squares method. Or a simpler algorithm can be used, while the form of $\mathbf{H}(\cdot)$ need not be known. For an input feature vector $\mathbf{f}_{ir}^{i}$, we compute its reconstruction coefficients $\mathbf{x}^{i}$ using its k-NNs in the infrared feature space. What we try to obtain is the reconstruction coefficient vector $\mathbf{y}^{i}$ which is used to reconstruct the visible feature $\mathbf{f}_v^{i}$ corresponding to $\mathbf{f}_{ir}^{i}$. We found the most similar coefficient vector $\mathbf{x}^{i'}$ with $\mathbf{x}^{i}$ in the infrared coefficient dataset, and we regard the corresponding visible coefficient vector $\mathbf{y}^{i'}$ as an estimate of $\mathbf{y}^{i}$. We call our method Sophisticated LLE.

5   Experimental Results

We use the public available database collected by Equinox Corporation [14] for our experiments. We select 70 subjects from the database, and each subject has 10 pairs of

visible and infrared images with different expressions. The long wave thermal infrared images are used because the points of image pairs are well-matched, even though they have lower resolution than the middle wave images. All the images have been manually registered to guarantee the eye centers and the mouth centers are well-registered. Some image pairs of the data set are shown in Fig. 2(a)(g).

We test our algorithm on the training set using the leave-one-out scheme, i.e. take one pair out from the database as the test images (the infrared image as the input and the visible image as the ground truth); all the pairs of the same subject are removed from the database as well; and the left pairs are taken as the training data.

(a)   (b)   (c)   (d)   (e)   (f)   (g)

Fig. 2. The results of face image conversion from thermal infrared images. (a)the input infrared image; (b) the result of our method using the prediction method of LLR using 5 canonical variates for each patch; (c) the result of our method using the prediction method of Sophisticated LLE; (d) the reconstruction face using 5 canonical features for each patch extracted from the ground truth; (e) the result using directly LLE method; (f) the result using the holistic method; (g) the ground truth.

There are several parameters to be chosen in our algorithm, such as the size of patches, the number of canonical variates k (the dimensionality of feature vector) we take for every patch, and the number of the neighbors we use to train the canonical directions.

Generally speaking, the correlation between pairs of infrared and visible patches of a smaller size is weaker, so the inference is less reasonable. While the larger size, makes the correlation stronger, but more canonical variates are needed to represent the patch, which makes training samples much sparser in the feature space. The size of images of our database is 110×86, and we choose the patch size of 9×9 with 3-px overlapping.

Since the projections (features) onto the former pairs of direction have stronger correlations, choosing fewer features makes the inference more robust, while choosing more features gives a more accurate representation of the original patch. Similarly, when we choose a larger number of neighbors, K, there are more samples, which makes the algorithm more robust but time-consuming. We choose 2~8 features and 30~100 neighbors for LLR, and we have the slightly different results.

We have compared our methods with other existing algorithms such as LLE and the holistic method. The results in Fig.2 show that our method is capable to preserve the global facial structure and to capture some facial detailed features such as wrinkles, mustache, and the boundary of nose. Our algorithm is also robust to facial expressions,
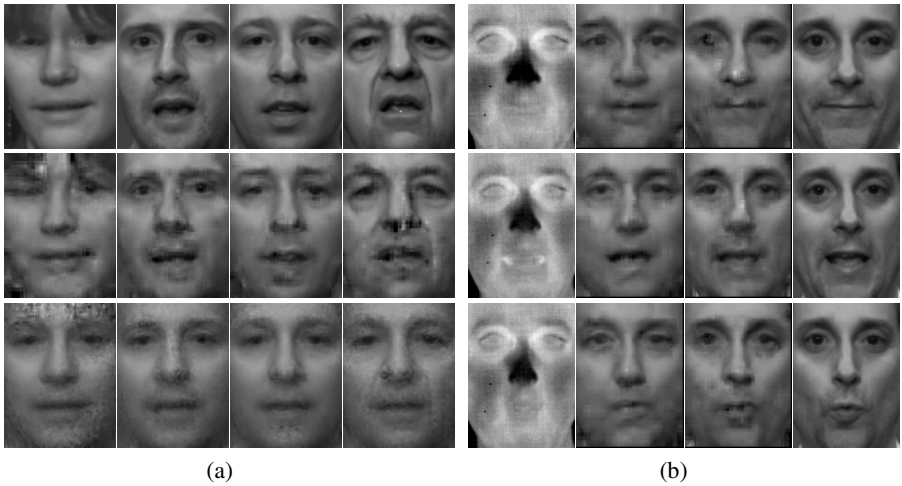

(a)   (b)

Fig. 3. (a) The comparison of reconstruction results using Eq(6) and Eq(7). The first row is the ground truth; the second row is the face reconstructed using Eq(7), and the third row using Eq(6). 5 canonical variates taken from the ground truth are used for each patch. It is clear that the reconstructions of Eq(7) contain more information than those of Eq(6). **(b)** The face image conversion results with different expressions of the same subject. The first column is the input infrared image; the second column is our conversion result; the third column is the reconstruction result using the canonical variates extracted from the ground truth; the last column is the ground truth.

as shown in Fig. 3 (**b**). The prediction methods proposed in Section 4.1 and 4.2 give slightly different results as shown in Fig. 2(b) (c).

Although our method is effective, there is still difference between our results and the ground truth. There should be two key points to account for it. First, the correspondence between the visible and the infrared images is a many-to-many mapping and infrared images contain less information than visible images. Second, our method tries to obtain the optimal result only in the statistical sense.

6 Conclusion and Future Work

In this paper we have developed an algorithm to render the visible facial images from thermal infrared images using canonical variates. Given an input thermal infrared image, we partition it into small patches, and for every patch we extract the CCA features. Then the features of the corresponding visible patch can be inferred by LLR or by sophisticated LLE, and the visible patch can be reconstructed by LLE using Eq(7) according to the inferred features.

We use CCA to extract features, which makes the correlation in the feature space are much stronger than that in the patch space. And using LLE to reconstruct the original patch from inferred features recovers some information lost in the infrared patch and in the feature-extraction process.

The experiments show that our algorithm is effective. Thought it cannot recover visible images the same as the ground truth because of less information of infrared images, but it does preserve some features of the ground truth such as the expression.

The future work includes: (1) applying the method in infrared face recognition to improve the recognition rate for it recovers some information lost in infrared images; (2) making the methods more robust to ill-registered images.

Acknowledgments

References

1. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.J.: Face Recognition: A Literature Survey. ACM Computing Surveys 35, 399–459 (2003)
2. Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A.: Recent Advances in Visual and Infrared face recognition—A Review. Computer Vision and Image Understanding 97, 103–135 (2005)
3. Bebis, G., Gyaourova, A., Singh, S., Pavlidis, I.: Face Recognition by Fusing Thermal Infrared and Visible Imagery. Image and Vision Computing 24, 727–742 (2006)

4. Heo, J., Kong, S.G., Abidi, B.R., Abidi, M.A.: Fusion of Visual and Thermal Signatures of with Eyeglass Removal for Robust Face Recognition. In: Proc. of CVPRW2004, vol. 8, pp. 122–127 (2004)
5. Baker, S., Kanade, T.: Limits on Super-Resolution and How to Break Them. IEEE Trans. On Pattern Analysis and Machine Intelligence 24, 1167–1183 (2002)
6. Chang, H., Yeung, D.Y., Xiong, Y.: Super-Resolution Through Neighbor Embedding. In: Proc. of CVPR2004, vol. 1, pp. 275–282 (2004)
7. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning Low-Level Vision. International Journal of Computer Vision 40, 25–47 (2000)
8. Cleveland, W.S., Devlin, S.J.: Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. Journal of the American Statistical Association 83(403), 596–610 (1988)
9. Shakhnarovich, G., Viola, P., Darrell, T.: Fast Pose Estimation with Parameter-Sensitive Hashing. In: Proc. of ICCV2003, vol. 2, pp. 750–757 (2003)
10. Roweis, S.T., Saul, L.K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290, 2323–2326 (2000)
11. Melzer, T., Reiter, M., Bischof, H.: Appearance Models Based on Kernel Canonical Correlation Analysis. Pattern Recognition 36, 1961–1971 (2003)
12. Reiter, M., Donner, R., Langs, G., Bischof, H.: 3D and Infrared Face Reconstruction from RGB data using Canonical Correlation Analysis. In: Proc. of ICPR2006, vol. 1, pp. 425–428 (2006)
13. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. on Pattern Analysis and Machine Intelligence 19, 711–720 (1997)
14. Socolinsky, D.A., Selinger, A.: A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery. In: Proc. of ICPR2002, vol. 4, pp. 217–222 (2002)
15. Weenink, D.: Canonical Correlation Analysis. In: IFA Proceedings, vol. 25, pp. 81–99 (2003)