Facial Image Reconstruction by SVDD-Based Pattern De-noising

Jooyoung Park[1,⋆], Daesung Kang[1], James T. Kwok[2], Sang-Woong Lee[3], Bon-Woo Hwang[3], and Seong-Whan Lee[3]

[1] Department of Control and Instrumentation Engineering, Korea University Jochiwon, Chungnam, 339-700, Korea

[2] Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

[3] Department of Computer Science and Engineering, Korea University, Anam-dong, Seongbuk-ku, Seoul 136-713, Korea

Abstract. The SVDD (support vector data description) is one of the most well-known one-class support vector learning methods, in which one tries the strategy of utilizing balls defined on the feature space in order to distinguish a set of normal data from all other possible abnormal objects. In this paper, we consider the problem of reconstructing facial images from the partially damaged ones, and propose to use the SVDD-based de-noising for the reconstruction. In the proposed method, we deal with the shape and texture information separately. We first solve the SVDD problem for the data belonging to the given prototype facial images, and model the data region for the normal faces as the ball resulting from the SVDD problem. Next, for each damaged input facial image, we project its feature vector onto the decision boundary of the SVDD ball so that it can be tailored enough to belong to the normal region. Finally, we obtain the image of the reconstructed face by obtaining the pre-image of the projection, and then further processing with its shape and texture information. The applicability of the proposed method is illustrated via some experiments dealing with damaged facial images.

1 Introduction

Recently, the support vector learning method has grown up as a viable tool in the area of intelligent systems. Among the important application areas for the support vector learning, we have the one-class classification problems [1, 2]. In the problems of one-class classification, we are in general given only the training data for the normal class, and after the training phase is finished, we are required to decide whether each test vector belongs to normal class or abnormal class. One of the most well-known support vector learning methods for the one-class problems is the SVDD (support vector data description) [1, 2]. In the SVDD, balls are used for expressing the region for the normal class. Since balls on the input domain can express only limited class of regions, the SVDD in general enhances its

⋆ Corresponding author.

D. Zhang and A.K. Jain (Eds.): ICB 2006, LNCS 3832, pp. 129–135, 2005.

expressing power by utilizing balls on the feature space instead of the balls on the input domain. In this paper, we extend the main idea of the SVDD for the reconstruction of partially damaged facial images [3]. Utilizing the morphable face model [4, 5, 6], the projection onto the spherical decision boundary of the SVDD, and a solver for the pre-image problem, we propose a new method for the problem of reconstructing facial images. The proposed method deals with the shape and texture information separately, and its main idea consists of the following steps: First, we solve the SVDD problem for the data belonging to the given prototype facial images, and model the data region for the normal faces as the ball resulting from the SVDD problem. Next, for each damaged input facial image, we perform de-noising by projecting its feature vector onto the spherical decision boundary on the feature space. Finally, we obtain the image of the reconstructed face by obtaining the pre-image of the projection with the strategy of [7], and further processing with its shape and texture information.

The remaining parts of this paper are organized as follows: In Section 2, preliminaries are provided regarding the SVDD, morphable face model, forward warping, and backward warping. Our main results on the facial image reconstruction by the SVDD-based learning are presented in Section 3. In Section 4, the applicability of the proposed method is illustrated via some experiments. Finally, in Section 5, concluding remarks are given.

2 Preliminaries

2.1 Support Vector Data Description

The SVDD method, which approximates the support of objects belonging to normal class, is derived as follows [1, 2]: Consider a ball B with the center $a \in \Re^d$ and the radius R, and the training data set D consisting of objects $x_i \in \Re^d$, $i = 1, \cdots, N$. Since the training data may be prone to noise, some part of the training data could be abnormal objects. The main idea of the SVDD is to find a ball that can achieve two conflicting goals simultaneously. First, it should be as small as possible, and with equal importance, it should contain as many training data as possible. Obviously, satisfactory balls satisfying these objectives can be obtained by solving the following optimization problem:

$$\begin{array}{ll} \min & L_0(R^2, a, \xi) = R^2 + C\sum_{i=1}^{N} \xi_i \\ \text{s. t.} & \|x_i - a\|^2 \leq R^2 + \xi_i, \ \xi_i \geq 0, \quad i = 1, \cdots, N. \end{array} \tag{1}$$

Here, the slack variable ξ_i represents the penalty associated with the deviation of the i-th training pattern outside the ball. The objective function of (1) consists of the two conflicting terms, *i.e.*, the square of radius, R^2, and the total penalty $\sum_{i=1}^{N} \xi_i$. The constant C controls relative importance of each term; thus called the trade-off constant. Note that the dual problem of (1) is:

$$\begin{array}{ll} \max_\alpha & \sum_{i=1}^{N} \alpha_i \langle x_i, x_i \rangle - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \langle x_i, x_j \rangle \\ \text{s. t.} & \sum_{i=1}^{N} \alpha_i = 1, \ \alpha_i \in [0, C], \quad \forall i. \end{array} \tag{2}$$

From the Kuhn-Tucker condition one can express the center of the SVDD ball as $a = \sum_{i=1}^{N} \alpha_i x_i$, and can compute the radius R utilizing the distance between a and any support vector x_i on the ball boundary. After the training phase is over, one may decide whether a given test point $x \in \Re^d$ belongs to the normal class utilizing the following criterion: $f(x) \stackrel{\triangle}{=} R^2 - \|x - a\|^2 \geq 0$. In order to express more complex decision regions in $\Re^d$, one can use the so-called feature map $\phi : \Re^d \rightarrow F$ and balls defined on the feature space F. Proceeding similarly as the above and utilizing the kernel trick $\langle \phi(x), \phi(z) \rangle = k(x, z)$, one can find the corresponding feature-space SVDD ball B_F in F, whose center and radius are a_F and R_F, respectively. If the Gaussian function $K(x, z) = \exp(-\|x - z\|^2/\sigma^2)$ is chosen for the kernel K, one has $K(x, x) = 1$ for each $x \in \Re^d$, which is assumed throughout this paper. Finally, note that in this case, the SVDD formulation is equivalent to

$$\begin{array}{rl} \min_\alpha & \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j K(x_i, x_j) \\ \text{s. t.} & \sum_{i=1}^{N} \alpha_i = 1, \ \alpha_i \in [0, C], \quad \forall i, \end{array} \tag{3}$$

and the resulting criterion for the normality is represented by

$$\begin{array}{l} f_F(x) \stackrel{\triangle}{=} R_F^2 - \|\phi(x) - a_F\|^2 \\ = R_F^2 - 1 + 2\sum_{i=1}^{N} \alpha_i k(x_i, x) - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j k(x_i, x_j) \geq 0. \end{array} \tag{4}$$

2.2 Morphable Face Model, Forward Warping and Backward Warping

Our reconstruction method is based on the morphable face model introduced by Beymer and Poggio [4], and developed further by Vetter *et al.* [5, 6]. Assuming that the pixelwise correspondence between facial images has already been established, a given facial image can be separated into the shape information and texture information. The two-dimensional shape information is coded as the displacement fields from a reference face, which plays the role of the origin in further information processing. On the other hand, the texture information is coded as an intensity map of the image which results from mapping the face onto the reference face. The shape of a facial image is represented by a vector $S = (d_1^x, d_1^y, \cdots, d_N^x, d_N^y)^T \in \Re^{2N}$, where N is the number of pixels in facial image, (d_k^x, d_k^y) the x, y displacement of a pixel that corresponds to a pixel x_k in the reference face and can be denoted by $S(x_k)$. The texture is represented as a vector $T = (i_1, \cdots, i_N)^T \in \Re^N$, where i_k is the intensity of a pixel that corresponds to a pixel x_k among N pixels in the reference face and can be denoted by $T(x_k)$. Before explaining our reconstruction procedure, we specify two types of warping processes: forward warping and backward warping. Forward warping warps a texture expressed in the reference face onto each input face by using its shape information. This process results in an input facial image. Backward warping warps an input facial image onto the reference face by using its shape information. This process yields a texture information expressed in reference shape. More details on the forward and backward warping can be found in reference [5].

3 Facial Image Reconstruction by SVDD-Based Learning

In the SVDD, the objective is to find the support of the normal objects, and anything outside the support is viewed as abnormal. On the feature space, the support is expressed by a reasonably small ball containing a reasonably large portion of the $\phi(x_i)$. A central idea of this paper is to utilize the ball-shaped support on the feature space for the purpose of correcting input facial images distorted by noises. More precisely, with the trade-off constant C set appropriately[1], we can find a region where the shape (or texture) data belonging to the normal facial images without noise generally reside. When a facial image (which was originally normal) is given as a test input x in a distorted form, the network resulting from the SVDD is supposed to judge that the distorted x does not belong to the normal class. The role of the SVDD has been conventionally up to this point, and the problem of curing the noise might be thought beyond the scope of the SVDD. However, here we observe that since the decision region of the SVDD is a simple ball B_F on the feature space F, it is quite easy to let the feature vector $\phi(x)$ of the distorted test input x move toward the center a_F of the ball B_F until it reaches the decision boundary so that it can be tailored enough to be counted normal. Of course, since the movement starts from the distorted feature $\phi(x)$, there are plenty of reasons to believe that the tailored feature $P\phi(x)$ still contain essential information about the original facial image. Thus, we claim that the tailored feature $P\phi(x)$ is the de-noised version of the feature vector $\phi(x)$. The above arguments together with additional step for finding the pre-image of $P\phi(x)$ comprise the essence of our method for facial image recovery. More precisely, our reconstruction procedure consists of the following steps:

1. Find the shape vectors $S_1, \cdots, S_N$ and texture vectors $T_1, \cdots, T_N$ for the given N prototype facial images.
2. Solve the SVDD problems for the shape and texture data belonging to the given prototype facial images, respectively, and model the data region for the shape and texture vectors of the normal faces as the balls resulting from the SVDD solutions, respectively.
3. For each damaged input facial image, perform the following:
   (a) Find the shape vector S of the damaged input facial image.
   (b) Perform de-noising for S by projecting its feature vector, $\phi_s(S)$, onto the spherical decision boundary of the SVDD ball on the feature space.
   (c) Estimate the shape of the recovered face, $\hat{S}$, by obtaining the pre-image of the projection $P\phi_s(S)$.
   (d) Find the texture vector T of the damaged input facial image.
   (e) Perform de-noising for T by projecting its feature vector, $\phi_t(T)$, onto the spherical decision boundary of the SVDD ball on the feature space.
   (f) Estimate the texture of the recovered face, $\hat{T}$, by obtaining the pre-image of the projection $P\phi_t(T)$.

[1] In our experiments, $C = 1/(N \times 0.2)$ was used for the purpose of de-noising.

(g) Synthesize a facial image for the reconstructed one by forward warping the estimated texture $\hat{T}$ with the estimated shape $\hat{S}$.

Steps 1, 3(a), and 3(d) are well explained in the previous studies of morphable face models [5, 8], and step 2 can be performed by the standard SVDD procedure. Steps 3(b)-(c) and 3(e)-(f) are carried out by the same mathematical procedure except that the shape about a pixel is a two-dimensional vector while the texture is one-dimensional. Therefore in the following description for steps 3(b)-(c) and 3(e)-(f), a universal notation is used for both S and T, *i.e.*, we will denote the object under consideration by $x \in \Re^d$, which can be interpreted as S or T according to which steps we are dealing with. Similarly, the feature maps for $\phi_s(\cdot)$ and $\phi_t(\cdot)$ are both denoted by $\phi(\cdot)$. As mentioned before, in step 2 of the proposed method, we solve the SVDD (3) for the shape (or texture) vectors of the prototype facial images $D \triangleq \{x_i \in \Re^d | i = 1, \cdots, N\}$. As a result, we find the optimal α_i along with a_F and R_F^2. In steps 3(b) and 3(e), we consider each damaged test pattern x. When the decision function f_F of (4) yields a nonnegative value for x, the test input is accepted normal as it is, and the de-noising process is bypassed. Otherwise, the test input x is considered to be abnormal and distorted by noise. To recover the de-noised pattern, an SVDD-based projection approach recently proposed by us [9] is used, in which we move the feature vector $\phi(x)$ toward the center a_F up to the point where it touches the ball B_F. Thus, the outcome of this movement is the following:

$$P\phi(x) = a_F + \frac{R_F}{\|\phi(x) - a_F\|}(\phi(x) - a_F). \tag{5}$$

Obviously, this movement is a kind of the projection, and can be interpreted as performing de-noising in the feature space. Note that as a result of the projection, we have the obvious result $\|P\phi(x) - a_F\| = R_F$. Also, note that with $\lambda \triangleq R_F / \|\phi(x) - a_F\|$, the equation (5) can be further simplified into

$$P\phi(x) = \lambda\phi(x) + (1 - \lambda)a_F, \tag{6}$$

where λ can be computed from

$$\lambda^2 = \frac{R_F^2}{\|\phi(x) - a_F\|^2} = \frac{R_F^2}{(1 - 2\sum_i \alpha_i K(x_i, x) + \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j))}. \tag{7}$$

In step 3(c) and 3(f), we try to find the pre-image of the de-noised feature $P\phi(x)$. If the inverse map $\phi^{-1} : F \rightarrow \Re^d$ is well-defined and available, this final step attempting to get the de-noised pattern via $\hat{x} = \phi^{-1}(P\phi(x))$ will be trivial. However, the exact pre-image typically does not exist [10]. Thus, we need to seek an approximate solution $\hat{x}$ instead. For this, we follow the strategy of [7], which uses a simple relationship between feature-space distance and input-space distance [11] together with the MDS (multi-dimensional scaling) [12]. After obtaining the de-noised vectors $\hat{S}$ and $\hat{T}$ from the above steps, we synthesize a facial image by forward warping the texture information $\hat{T}$ onto the input face by using the shape information $\hat{S}$. This final synthesis step is well explained in [5, 8].

4 Experiments

For illustration of the proposed method, we used two-dimensional images of Caucasian faces that were rendered from a database of three-dimensional head models recorded with a laser scanner (*Cyberware*^{TM}) [5, 6]. The resolution of the images was 256 by 256 pixels, and the color images were converted to 8-bit gray level images. Out of the 200 facial images, 100 images were randomly chosen as the prototypes for the SVDD training (step 2), and the other images were used for testing our method. For the test data set, some part of each test image was damaged with random noises. When extracting the S and T information from the damaged test input images, manual intervention based on the method of [13] was additionally employed. The first row of Fig. 1 shows the examples of the damaged facial images. The second and third row of Fig. 1 show the facial images reconstructed by the proposed method and the original facial images, respectively. From the figure we see that most of the reconstructed images are similar to the original ones.

Fig. 1. Examples of facial images reconstructed from the partially damaged ones. The images on the top row are the damaged facial images, and those on the middle row are the facial images reconstructed by the proposed method. Those on the bottom row are the original face images.

5 Concluding Remarks

In this paper, we addressed the problem of reconstructing facial images from partially damaged ones. Our reconstruction method depends on the separation of facial images into the shape vectors S and texture vectors T, the SVDD-based de-noising for each of S and T, and finally the synthesis of facial images from the de-noised shape and texture information. In the SVDD-based de-noising, we utilized the SVDD learning, the projection onto the SVDD balls in the feature space, and a method for finding the pre-image of the projection. Experimental results show that reconstructed facial images are natural and plausible like original facial images. Works yet to be done include extensive comparative studies, which will reveal the strength and weakness of the proposed method, and further use of the proposed reconstruction method to improve the performance of face recognition systems.

Acknowledgments

We would like to thank the Max-Planck-Institute for providing the MPI Face Database.

References

1. D. Tax and R. Duin, "Support Vector Domain Description," *Pattern Recognition Letters*, vol. 20, pp. 1191–1199, 1999.
2. D. Tax, *One-Class Classification, Ph.D. Thesis*, Delft University of Technology, 2001.
3. B.-W. Hwang and S.-W. Lee, "Reconstruction of partially damaged face images based on a morphable face model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 365-372, 2003.
4. D. Beymer and T. Poggio, "Image representation for visual learning," *Science*, vol. 272, pp. 1905-1909, 1996.
5. T. Vetter and N. E. Troje, "Separation of texture and shape in images of faces for image coding and synthesis," *Journal of the Optical Society of America A*, vol. 14, pp. 2152-2161, 1997.
6. V. Blanz, S. Romdhani, and T. Vetter, "Face identification across different poses and illuminations with a 3d morphable model," *Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition*, Washington, D.C., pp. 202-207, 2002.
7. J. T. Kwok and I. W. Tsang, "The pre-image problem in kernel methods," *IEEE Transactions on Neural Networks*, vol. 15, pp. 1517–1525, 2004.
8. M. J. Jones, P. Sinha, T. Vetter, and T. Poggio, "Top-down learning of low-level vision task[Brief Communication]," *Current Biology*, vol. 7, pp. 991-994, 1997.
9. J. Park, D. Kang, J. Kim, I. W. Tsang, and J. T. Kwok, "Pattern de-noising based on support vector data description," To appear in Proceedings of International Joint Conference on Neural Networks, 2005.
10. S. Mika, B. Schölkopf, A. Smola, K. R. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature space," *Advances in Neural Information Processing Systems*, vol. 11, pp. 536–542, Cambridge, MA: MIT Press, 1999.
11. C. K. I. Williams, "On a connection between kernel PCA and metric multidimensional scaling," *Machine Learning*, vol. 46, pp. 11–19, 2002.
12. T. F. Cox and M. A. A. Cox, "Multidimensional Scaling," *Monographs on Statistics and Applied Probability*, vol. 88, *2nd Ed.*, London, U.K.: Chapman & Hall, 2001.
13. B.-W. Hwang, V. Blanz, T. Vetter, H.-H. Song and S.-W. Lee, "Face Reconstruction Using a Small Set of Feature Points," *Lecture Notes in Computer Science*, vol. 1811, pp. 308-315, 2000.