Normalization of Face Images using Few Anchors

Daniel Reisfeld Nur Arad and Yehezkel Yeshurun
School of Mathematical Sciences
Tel Aviv University, Israel
reisfeld@math.tau.ac.il, nur@math.tau.ac.il, hezy@math.tau.ac.il

Abstract

We present a procedure for facial normalization which is determined by the translocation of few anchors in two sets. The first set determines the affine part of the transformation and is used for overcoming change in viewpoint and scale. The second set is used for normalization of expression. A generalization of radial basis functions theory is used for generating the appropriate mapping. Only a few $(1-3)$ anchors were used to demonstrate elaborate facial expressions without any face model.

1 Introduction

Recognition of faces is different from most other recognition tasks since faces undergo major spatial deformations due to changes of viewpoint and of expressions, while small changes may distinguish between persons. We presented [10] a three step strategy for the construction of face recognition systems: Locating the face and some facial features in the image; transforming the face into a normal form; and applying a general purpose classifier. We suggested the location of facial features using the generalized symmetry transform [11, 12, 13, 10], and used various general purpose classifiers [5, 10]. In this paper we concentrate on the normalization procedure. It is clear that virtually every face recognition scheme can also benefit from an image warp technique that can transform facial images into a standard form, which reduces the differences between various images of the same individual provided that the differences between different ones are preserved.

The difficulty of locating minute facial features precludes the use of complex face models such as [15, 8, 9]. An alternative is to use a general purpose two dimensional approach. This is justified from the neurobiological and psychophysical viewpoints (as reviewed in [10]). Affine mappings can be specified by the translocation of three anchor points, and be used to compensate for changes in viewpoint. A special case are the similarity mappings – rotation, translation and scaling, which are determined by two anchors. The use of the location of the eyes for face normalization dates back to 1878 when Galton superimposed facial images using a scaling transformation [6]. A natural generalization of Galton's technique is to use general affine mappings. We have successfully used the affine mapping determined by the locations of the eyes and mouth for the purpose of face recognition [5, 13, 7, 10].

Locating the centers of the eyes and mouth enables the use of an affine transformation. We define a transformation based on more anchor points, generalizing the affine transformation and compensating for changes in facial expression. Of crucial importance is the use of a small number of anchors, since the automatic extraction of many anchors is probably less plausible than face recognition with affine normalization. The anchors should be allowed to lie in general positions (rather than on a regular grid) in order to coincide with the position of the facial features.

2 The Mapping

The theory of radial basis functions has proven to be an effective tool in multivariate interpolation problems of scattered data. It also provides an attractive framework for image warping. Bookstein [3] suggested the use of a subclass of these functions, known as *thin-plate splines*, for image registration purposes since their use minimizes a global measure of warp energy. However, global effects are in many cases undesirable – many objects undergo elaborate transformations which have local, as well as global, components. We introduced radial based transformations that may be used to emphasize a local influence of the anchor points [1, 2]. Other global constraints are incorporated as well, and tradeoff parameters govern the effect of the various constraints.

Given a univariate function $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ one may attempt to interpolate the scattered d-dimensional data $(\bar{x}^i, F_i)$, $\bar{x}^i \in \mathbb{R}^d$, $F_i \in \mathbb{R}$, $i = 1, 2, \ldots, N$, using a *pure radial sum*:

$$S(\bar{x}) = \sum_{i=1}^{N} a_i g\left(\|\bar{x} - \bar{x}^i\|\right)$$

where $\|\cdot\|$ denotes the usual Euclidean norm on $\mathbb{R}^d$. The choice of a radial function reflects the fact that the scattered data has no preferred orientation, and the fact that for given i the data point $\bar{x}^i$ equally effects all points of equal distance to $\bar{x}^i$. Interpolation by

1051-4651/94 $04.00 © 1994 IEEE

a pure radial sum is possible whenever the system of linear equations

$$G\bar{a} = F$$

has a unique solution, where $G = g_{ij} = g\left(\left\|\bar{x}^i - \bar{x}^j\right\|\right), \bar{a} = (a_1, a_2, \ldots, a_N)^T, F = (F_1, F_2, \ldots, F_N)^T$. Some classes of functions for which a unique solution exists for any N distinct points $\bar{x}^i \in \mathbb{R}^d$ are found in [4].

A generalization of this theory to functions $T : \mathbb{R}^2 \to \mathbb{R}^2$ is interpolation of the form $T(\bar{x_i}) = \bar{y_i}$ for $i = 1, 2, \ldots, N$, where $\bar{x_i}, \bar{y_i} \in \mathbb{R}^2$. We define $T : \mathbb{R}^2 \to \mathbb{R}^2$, the two-dimensional radial basis function transformation as

$$T(\bar{x}) = A(\bar{x}) + R(\bar{x})$$

where $A(\bar{x}) = M\bar{x} + \bar{b}$ is a 2D affine transformation (M is a 2×2 real matrix), and $R(\bar{x})$ is a radial transformation defined by $R(\bar{x}) = (R_X(\bar{x}), R_Y(\bar{x}))$ R_X and R_Y are both pure radial sums. This definition enables the handling of the facial expression once the affine component (change of viewpoint) is determined.

Controlling the affine component, $A(\bar{x})$, is carried out according to the number of anchor points in an *affine set*. If no points are specified, the affine component is the identity mapping; if one point is specified - translation; two points - rotation, translation and scaling; three points - general affine transformation; more than three - general affine transformation determined by a least-square approximation procedure. Alternatively, one can decide that the most general affine mapping admitted is similarity, in which case a least-square approximation scheme is needed when the affine set consists of three or more points.

After the affine component, $A(\bar{x})$, has been computed, we turn to determining the radial component, $R(\bar{x})$. From the above equations it follows that the radial component satisfies

$$R(\bar{x_i}) = \bar{y_i} - A(\bar{x_i}) \quad , \quad i = 1, 2, \ldots, N.$$

Since the affine component is determined beforehand, the last equation is equivalent to an interpolation problem with pure radial sums, i.e. the coefficients a_i in each of the coordinates (i.e. $k = 1, 2$) are determined by solving this linear system

$$\sum_{j=1}^{N} a_{j,k} g\left(\|\bar{x}_i - \bar{x}_j\|\right) = y_{i,k} - (A(\bar{x}_i))_k \quad , \quad i = 1, \ldots, N$$

The theory of radial basis functions ensures the solvability of this linear system provided the anchor points are distinct. However, the radial basis function, g, must fulfill certain conditions. We introduced [2] the transition function:

$$g_\sigma(t) = \begin{cases} 1 - (t/\sigma)^2(3 - 2t/\sigma) & 0 \le t \le \sigma \\ 0 & t > \sigma \end{cases}$$

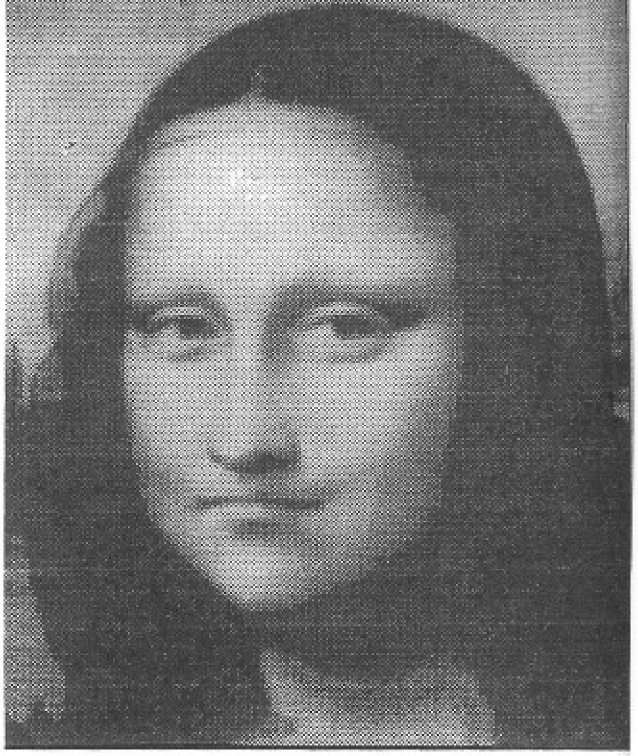

Figure 1: Transformed Monalize using onle one anchor point (on the mouth)

as a new radial function. This function has local support and ensures solvability and stability in all the practical cases (as discussed in length in [2]). It is superior to the known radial functions for our application since it has local support, and thus enables local influence of the anchors.

3 Demonstrations

We implemented the mapping with a comparable run-time to that of loading an image and presenting it on the screen. In order to achieve this fast implementation, some special care should be taken at certain points as discussed in [1].

The power of the decomposition of the affine and elastic components of the transformation is best demonstrated by the application of the transformation using only *one* anchor point. Since no points are specified for the affine component, it is taken as unity.

Figure 2 demonstrates the use of the mapping for normalizing facial expressions using a small number of anchors. a and b are snapshots of the same individual. The difference in the expression is natural and not due to any artificial transformation. The two images are already aligned using the similarity transformation determined by the eyes' location. The location of the

Figure 2: Facial expression normalization: Left and middle – two original images with three anchors marked on them. Right – the warped middle image using the average anchors (of left and middle images) as target.

center of the mouth is not easily determined in this case, and the general affine transformation, using the location of the mouth as well, is neither superior nor inferior for normalization purposes to the similarity transform. A second stage involves compensating for the deformation caused by change in expression. We now set the affine part to unity and look for proper anchors which determine the elastic component of the mapping. Recall that this process is equivalent to simultaneously fixing the affine component as the similarity transform determined by the location of the eyes and determining the elastic part.

References

[1] N. Arad, N. Dyn, D. Reisfeld, and Y. Yeshurun. Image warping by radial basis functions; application to facial expressions. *Journal of Computer Vision, Graphics, and Image Processing*, 56(2):161–172, March 1994.

[2] N. Arad and D. Reisfeld. Image warping using few anchor points and radial functions. *Computer Graphics Forum*, 1994. in press.

[3] F.L. Bookstein. Principal warps: Thin plate splines and the decomposition of deformations. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 11:567–585, 1989.

[4] N. Dyn. Interpolation and approximation by radial and related function. In C.K. Chui, L.L. Schumaker, and J.D. Ward, editors, *Approximation Theory VI*, volume 1, pages 211–234. Academic Press, 1989.

[5] S. Edelman, D. Reisfeld, and Y. Yeshurun. Learning to recognize faces from examples. In *Proceedings of the 2nd European Conference on Computer Vision*, pages 787–791, Santa Margherita Ligure, Italy, May 1992.

[6] F. Galton. Composite portraits, made by combining those of many different persons, into a single, resultant figure. *journal of the Anthropological Institute*, 8:132–144, 1879.

[7] N. Intrator, D. Reisfeld, and Y. Yeshurun. Face recognition using a hybrid supervised/unsupervised neural network. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, Jerusalem, Israel, 1994.

[8] N. Magnenat-Thalmann, P. Primeau, and D. Thalmann. Abstract muscle action procedures for face animation. *The Visual Computer*, 3:290–297, 1988.

[9] S. Platt and N.I. Badler. Animating facial expression. *Computer Graphics*, 15(3):245–252, 1981.

[10] D. Reisfeld. *Generalized symmetry transforms: attentional mechanisms and face recognition*. PhD thesis, Computer Science Department, Tel-Aviv University, January 1994.

[11] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Detection of interest points using symmetry. In *Proceedings of the 3rd International Conference on Computer Vision*, pages 62–65, Osaka, Japan, December 1990.

[12] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context free attentional operators: the generalized symmetry transform. *International Journal of Computer Vision*, 1994. special issue on qualitative vision.

[13] D. Reisfeld and Y. Yeshurun. Robust detection of facial features by generalized symmetry. In *Proceedings of the 11th IAPR International Conference on Pattern Recognition*, pages A117–120, The Hague, The Netherlands, August 1992.

[14] D. Reisfeld, Y. Yeshurun, N. Intrator, and S. Edelman. Efficient recognition of automatically normalized face images following random and epp-based dimensionality reduction. 1993.

[15] K. Waters and D. Terzopoulos. Modeling and animating faces using scanned data. *The journal Of Visualization and Computer Animation*, 2(4):123–128, 1991.