Towards a Supervised Incremental Learning System for Automatic Recognition of the Skeletal Age

Fernando Montoya Manzano[1], Salvador E. Ayala-Raggi[1(✉)], Susana Sánchez-Urrieta[1], Aldrin Barreto-Flores[1], José Francisco Portillo-Robledo[1], and Verónica Edith Bautista-López[2]

[1] Facultad de Ciencias de la Electrónica, Benemérita Universidad Autónoma de Puebla, Avenue San Claudio and 18 Sur, Col. Jardines de San Manuel, 72570 Puebla, Puebla, Mexico
fmm_ferix@hotmail.com, {saraggi,surrieta,abarreto,portillo}@ece.buap.mx

[2] Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, Avenue San Claudio and 18 Sur, Col. Jardines de San Manuel, 72570 Puebla, Puebla, Mexico

Abstract. In this work, we proposed and developed a simple system to estimate skeletal maturity based in using Active Appearance Models in order to create an increasing set of shape-aligned training images which are incrementally stored and used by a $K-NN$ regression classifier. For that purpose, we designed an original layout of landmarks to be located in representative regions of the radiographical image of the hand. Our results show that is possible to use pixels directly as classification features as long as the training and testing images have been previously aligned in shape and pose.

Keywords: Skeletal maturity recognition · Bone age estimation · Active appearance models · $K-NN$ regression

1 Introduction

Skeletal maturity estimation is an important issue in the proper diagnosis of a great set of diseases and growth problems. Estimation is often performed manually by the radiologist who subjectively compares bones and joints between the test image, normally a radiographic image of the left hand, and several template images taken from a standard handbook [1,2]. This procedure is prone to produce wrong or inaccurate age classification mainly caused by two important reasons: human errors due to subjective comparisons, and the fact that there is a

V.E. Bautista-López—The authors wish to thank Imagen Exakta S.A. de C.V., Dr. Patricia Ayala-Raggi MD, and Dr. Juan de Dios Meza-Rojas MD for providing them with a special permission to use a set of anonymous radiological images for academic purposes.

© Springer International Publishing Switzerland 2016
J.F. Martínez-Trinidad et al. (Eds.): MCPR 2016, LNCS 9703, pp. 346–355, 2016.
DOI: 10.1007/978-3-319-39393-3_34

single template image per age in the manual. Even though most utilized manuals have been carefully designed by experts in the field, the fact of using a single prototype hand image could be unsuitable because the great appearance variability in human hands. On the other hand, other approaches commonly used by some radiologists based in calculating different scores for different groups of bones and then carrying out a weighted average is extremely impractical for many physicians. And although it is more accurate than the former methods, it is subjective too [3,4]. Many automatic approaches have been proposed to overcome the problem of subjectivity. Niemeijer et al. [5] proposed an automatic system based on Tanner-Whitehouse method. By using a large number of Regions Of Interest, or shortly ROIs, labeled with ages by a radiologist, a mean image for each age is computed. Then, when a new input ROI is entered for classification, an active shape model, or ASM [6], is used to align that to the mean images, and by using correlation the algorithm selects the age label assigned to the mean image that best matches the input ROI. In that approach we observe two drawbacks: first, the mean images used do not contain the variability observed in the original images utilized to construct them. Second, the set of ages is finite, and therefore the estimated age is not a continuous variable. In [3], Hsieh et al. proposed to manually extract geometric features from carpal bones radiographs for ages from one to eight years, and then use artificial neural networks, shortly ANNs. Somehow, it is in part of an expert system combined with a supervised learning approach. The purpose of our work was not to automate some of the known clinical methods for bone age assessment [1,2], as done in some previous works [7]. Instead of that, we found it very interesting to investigate and design an automatic bone age learning system that is capable to learn from test examples. As previous knowledge, only the age labels from those examples should be used. Although there have been efforts to achieve bone age estimation algorithms based only on statistically learned features from lots of examples like [8,9], we have seen that no efforts have been done in designing automatic learning systems for estimating bone age with the capability of improving their performance and learning as more images are presented to the system. It would be practical and useful that a system could improve its performance even when it is being used for testing. Typical classification algorithms that are inherently incremental in that sense are precisely those based on the K nearest neighbors search or simply $K-NN$ algorithms. In our work, we have used a regression version of the classical $K-NN$ algorithm which is based on radial basis functions. At first glance, it would seem possible to add new images to the system every time you want to increase learning, and therefore the performance in the classification process too. However, and considering that the orinal images are not aligned, it would be necessary a considerably large number of training images in order to reach an acceptable classification rate. To address this problem, we have implemented a previous alignment algorithm based on a pre-trained Active Appearance Model, or shortly AAM [10], which segments the region of the hand and then warps its texture into a normalized hand shape. Thus, each example to add will become shape-aligned, and important features will be aligned too. This shape-normalized

image data and the reduced set of shape parameters returned by the AAM fitting process can be joined together to create a new feature vector which can be added to a knowledge base or training examples set where a $k-NN$ algorithm performs classification.

2 System Overview

In our proposed recognition system, we distinguish two important moments, the first one was the development stage and the second one is the usage stage. During the development an AAM model [11] was created from a set of manual labeled hand images with sufficient variability in shape and age. The purpose of this AAM model is to align or fit the model to a new test hand image during the usage stage. Then, during the usage stage, the user has two options: test a new image, or enter a new image for incremental learning. In both options, the user enters a new image to the system and the AAM alignment is carried out after the user provides an initial location to the model. When the fitting process ends and the model converges and resembles the original hand image, we can segment the hand's region and warp the original texture from the test image into a shape-normalized grid in order to create a shape-normalized hand image. This image is joined to the shape parameters returned by the AAM fitting process creating a feature vector. If the user selected the *test* option, the feature vector can be classified using $K-NN$ regression. On the other hand, if the user selected the *incremental learning* option, the feature vector together with the age label is added to a knowledge base, which is a set of these feature vectors. At the starting point of the usage stage, when no test images have been given to the system, the knowledge base is set up with the feature vectors corresponding to the hand images used to train the AAM model. Therefore the system can learn during the usage stage and it is not required to re-train the AAM model, which would be an expensive operation. Figure 1 shows the whole process.

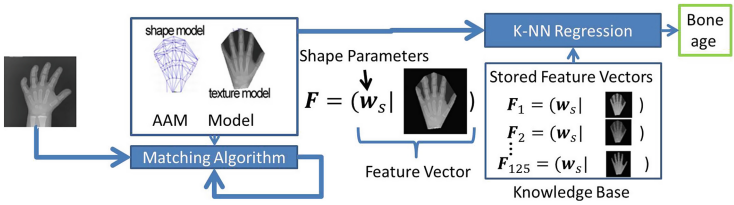

Fig. 1. Overall process diagram

3 Creating an Active Appearance Model of Radiographical Hand Images

We have built an AAM model in order to model the shape and texture variability of hands from radiographical images. $AAMs$ are parametric models based

in applying principal component analysis, or PCA, both to textures and shapes from a large set of training hand images. Shape is modeled by manually placing a set of landmarks over distinctive points on each hand image. We call a *shape* or simply $\mathbf{s}_i$ to the set of landmarks (x_j, y_j) of a particular hand image. Thus, and after an aligning procedure, we apply PCA to the set of $\mathbf{s}_i$ in order to obtain a reduced set of *eigenshapes* capable to represent a high percentage of the shape variance showed by the training set. On the other hand, textures contained inside the $\mathbf{s}_i$ are normalized in shape by mapping o warping them to the mean shape before applying PCA to them. Similarly to the shape, we will obtain a reduced set of *eigentextures*, [12], capable to represent a high percentage of the texture variance showed by the training set. This capability of high representativeness using just a reduced set of *eigenshapes* and *eigentextures* is possible thanks to the high similarity among the training images. By computing linear combinations of the *eigenshapes* and *eigentextures* we are able to reconstruct every hand of the training set, or even create or synthesize a novel hand image. The coefficients, more commonly known as *weights*, of those linear combinations are the parameters of the AAM model. Once the AAM model has been created, it can be used for aligning it to an input image. By using a fitting iterative algorithm similar to that implemented in [10], it is possible to recover the shape and texture parameters for the model which best matches the hand portrayed in the input image.

In order to construct the shape model, we designed a proper layout of landmarks and a respective grid of triangles such that the triangles geometrically could never flip due to variations in shape present in the training set. Figure 2 shows our proposed layout of landmarks placed in the corners of the hand bones, and Fig. 3 illustrates our proposed manual triangulation used in our project compared with an automatic *delaunay* triangulation [13] automatically generated by MATLAB. We can see that our triangulation is more robust to variation in landmark locations than the *delaunay* triangulation because the landmarks of some triangles are near to be collinear.

Fig. 2. Set of landmarks placed in distinctive points of the bones structures

Once the images have been manually labeled with the same set of landmarks, an iterative procedure based in Procrustes Analysis [11,14] has to be applied to the set of *shapes* in order to align them in the rigid body sense, i.e. *scale*,*rotation*, and *traslation*. After applying PCA to the that set of aligned shapes, and giving a proper set of shape parameters, every *shape* in the training

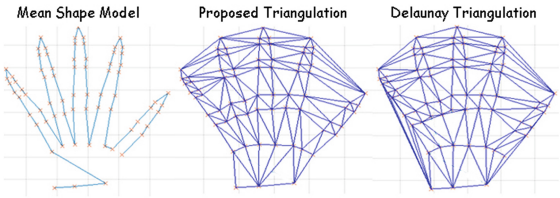

Fig. 3. Mean shape model, proposed triangulation and triangulation calculated by *delaunay* method

set can be reconstructed by the following expression,

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{Q}_s \mathbf{w}_s \tag{1}$$

where $\mathbf{s}$ is the output shape, $\bar{\mathbf{s}}$ is the mean shape, $\mathbf{Q}_s$ is a matrix whose columns are *eigenvectors* that we call *eigenshapes*, and $\mathbf{w}_s$ is the vector of shape parameters. The number of *eigenshapes* and therefore the number of shape parameters can be dramatically reduced if we preserve only the *eigenshapes* corresponding to their largest respective *eigenvalues*. Thus, the number of *eigenshapes* needed to represent a high percentage of the variance in the training set could be much smaller than the number of training examples. By using Eq. 2 and the respective set of *shape* parameters corresponding to each training image, textures of the training images can be warped into the mean shape, creating a set of $shape - normalized$ textures. Figure 4 illustrates that process. Thus, we can apply PCA to this new set of $shape - normalized$ training images. This process is known as $eigenfaces$, see [12].

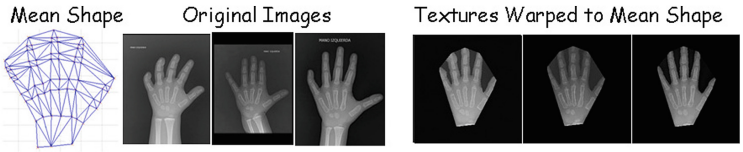

Fig. 4. Example of how warped images to mean shape look like.

Any texture from the training set can be approximated using the following expression:

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{w}_g \tag{2}$$

where $\mathbf{g}$ is the reconstructed texture, $\bar{\mathbf{g}}$ is the mean texture, $\mathbf{Q}_g$ is a matrix whose columns we call *eigentextures* and $\mathbf{w}_g$ is a vector of texture weights or texture parameters. The number of *eigentextures* and therefore the number of texture parameters can be reduced if we preserve only the *eigentextures* corresponding to their largest respective *eigenvalues*. The AAM model parameters are completed when pose parameters (s, Θ, t_x, t_y) (scaling, in-plane rotation angle, translation) are added, $\mathbf{p} = (\mathbf{w}_s^T, \mathbf{w}_g^T | s, \Theta, t_x, t_y)$.

3.1 Fitting the *AAM* Model to an Input Hand Image

We have used the iterative algorithm described in [10] to align the *AAM* model to an input hand image. Although in essence it is the same alignment algorithm, some modifications have been necessary in order to work properly with radiographic images instead of face images. In face images the background is independent from face. Faces appear like solid objects over the background. However, in radiographic images, hands are objects with certain degree of transparency over a flat background. Therefore, the gray level of the background is present inside the hand structure.

The alignment process of the model to an input image is computed by using an iterative algorithm (see [10] for details) which calculates the residual $\mathbf{r}(\mathbf{p})$ (the current difference between the model and image) each iteration. This residual is always measured in a shape-normalized texture frame). By assuming the relationship between the residual $\mathbf{r}(\mathbf{p}+\Delta\mathbf{p})$ and $\Delta\mathbf{p}$ as approximately constant (denoted as $\frac{\delta r}{\delta p}$), a proper additive increment to the parameters $\Delta\mathbf{p}$ can be computed in each iteration by

$$\delta\mathbf{p} = -\mathbf{R}\mathbf{r}(\mathbf{p}) \quad where \quad \mathbf{R} = (\frac{\delta r}{\delta \mathbf{p}}^T \frac{\delta r}{\delta \mathbf{p}})^{-1} \frac{\delta r}{\delta \mathbf{p}}^T \tag{3}$$

where $\frac{\delta r}{\delta \mathbf{p}}$ is a Jacobian matrix composed whose number of columns equals the number of model parameters. The jth column of this Jacobian was computed by systematically displacing each parameter from its initial value for the synthetic mean model. In order to avoid numeric errors and for increasing speed, the pseudo-inverse matrix of Moore-Penrose is used instead of normal matrix inverse during the calculation of Matrix $\mathbf{R}$.

In [10], the *AAM* alignment algorithm works fine on an input image with black background if and only if the matrix $\mathbf{R}$ has been computed from synthetic face images with black background. Similarly, in the case of aligning to radiographical images, a proper alignment will be posible if and only if the matrix $\mathbf{R}$ has been computed from synthetic hand images with a background which depends on the gray level inside the hand structure. In order to fulfill the former requirement, a synthetic background gray level has been generated during the calculation of the matrix $\mathbf{R}$. After multiple tests, we proposed a method to compute the background gray level by

$$b = \bar{m}_{x,y} - 0.8\sigma_{x,y} \tag{4}$$

where $\bar{m}_{x,y}$ is the mean of the gray level of all pixels (x, y) inside the synthetic hand region (including inter-fingers spaces), and $\sigma_{x,y}$ is the standard deviation of the gray level for all same pixels (x, y).

Figure 5 shows two cases in the process of construction of the matrix $\mathbf{R}$. In both cases mean texture $g(\mathbf{p} = 0)$ was altered by systematically displacing parameters $\mathbf{p}$ in an increment $\Delta\mathbf{p}$. At the left, the background gray level was set to zero. At the right, the background gray level was set to the value calculated by Eq. 4.

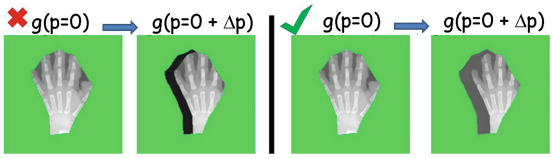

Fig. 5. Two cases in the construction of matrix **R**. Left: In this case the background in synthetic images is set to zero, and the residuals calculated for construction of the matrix **R** are different in nature to those obtained during the fitting process producing a wrong alignment. Right: The background in synthetic images is calculated from the hand internal pixels, and the residuals calculated for construction of the matrix **R** are similar to those computed during the fitting process

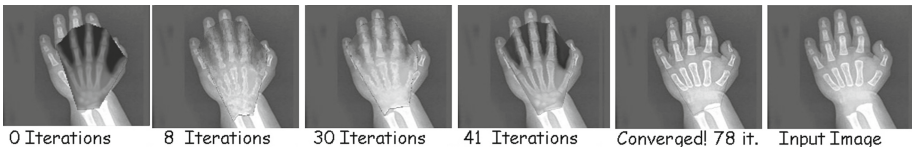

Fig. 6. Iterative alignment of the *AAM* model to an input image.eps

We found that by using Eq. 4 the convergence of the fitting process was reached correctly. Figure 6 illustrates the alignment of fitting process on a baby radiographic hand image.

4 Bone Age Classification

The shape parameters returned from fitting the model to a new image are joined to the shape-normalized texture from the original image in order to create a feature vector useful for $K - NN$ classification.

In theory, we can use the parameters of shape and texture returned by the *AAM* alignment process to create short feature vectors which are suitable for classification. However, proceeding in that way does not allow us to implement an incremental learning process because a new *AAM* model would have to be rebuilt every time a new example image is added to our knowledge base. Of course, this is an expensive operation. Therefore, in order to design a practical incremental learning system, we have chosen to join the shape parameters returned from the fitting process with the shape-normalized texture from the input image, creating new vector of aligned features $\mathbf{F} = (\mathbf{w}_s^T | \mathbf{g}^T)$.

Thus, we do not need to recompute the *AAM* model every time a new image is added for learning. Therefore, we must ensure that *AAM* model is made from enough variability in shape and texture in order to be capable to fit almost every input image.

4.1 Age Classification by $K - NN$ Regression

The estimated feature vector $\mathbf{F}$ can be classified by comparing it with those stored in a set that we call *knowledge base*. In order to carry out the classification we implemented a K-Nearest Neighbor regression algorithm [15,16]. This process consists of finding the k feature vectors stored in the knowledge base which are the nearest in the Euclidean distance sense. We use these distances d_i $(i = 1, ..., k)$ to compute respective weight values W_i for each neighbor such that the greater the distance the weight value will be lower.

$$W_i = \exp \frac{-d_i^2}{2\sigma^2} \tag{5}$$

where σ is a constant that can be obtained by trial and error. Therefore, bone age can be computed as a weighted average of the respective age labels of the k neighbors

$$\mathbf{age} = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \tag{6}$$

where Y_i are the age values of the k neighbors.

5 Experimental Results

To test our age recognition method, we used an image set composed by 165 radiographical images all them cropped and resized to 256 $\times$ 256. 125 images from that set were utilized for training and learning, and the remaining 40 were reserved for testing. From the training and learning set, 65 images were used for training the AAM model. The shape-normalized textures of these 65 images joined to their respective shape parameters were used as an inicial set of feature vectors for the knowledge base. The remaining 60 were used for incremental learning during the usage stage.

For the AAM model we designed a layout of 71 landmarks located over distinctive corners of hand bones. We decided to preserve 30 *eigenshapes* which provide us 99 % of the variance observed in the training set. Similarly, we used 25 *eigentextures* representing 99 % of the variance observed in the training set. For $K - NN$ regression, by setting $k = 5$ we obtained the best results. Figure 7(B) shows the estimated ages for 40 tests by using a knowledge base containing 125 feature vectors. We obtained a Mean Absolute Error (MAE) of 1.8 years, a Mean Error (ME) of -0.08 years, and a Root Mean Square Error ($RMSE$) of 1.87 years.

In order to evaluate the capability of our system to incrementally learn, we tested age recognition using the set of 40 test images every time the knowledge base was incremented with additional subsets of 10 feature vectors. The process started with a knowledge base containing just the first 65 features vectors, and finished with 125 feature vectors. Figure 7(A) shows a gradual decrease in the error rate ($R - MSE$) every time that the knowledge base is incremented.

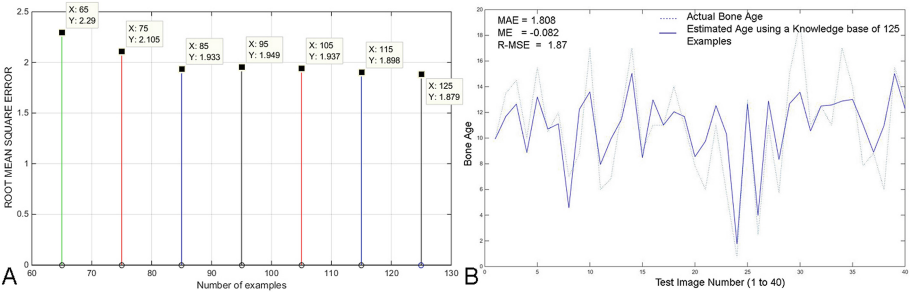

Fig. 7. (A). The root mean square error was measured for 40 test images each time a set of 10 learning images was added to the knowledge base. Learning images have been added from image 65 to image 125. (B). Actual ages vs estimated ages with 125 examples

6 Conclusions and Future Work

In this paper we have proposed and implemented an automatic skeletal maturity recognition system which is capable of estimating bone age with an acceptable accuracy. In contrast to other works, we proposed to investigate the problem of bone age estimation in the context of supervised learning, based in giving only the training examples and their age labels to the computer. Just images and ages should be used as previous knowledge. In addition, we proposed a simple incremental learning methodology which gradually reduces the classification error when more images are presented to the system during the usage stage, without the need to re-train the AAM model. This incremental learning skill can not be implemented using only the AAM parameters, because the addition of novel training images implies to make a reconstruction of the AAM model, and that is not practical. On the other hand, unaligned images can be directly used in classifiers like $k - NN$ or neural networks but the required quantity of them should be really large in order to obtain acceptable results. In our approach, we proposed to use an AAM model just for segmenting and aligning the hand region in order to produce a shape-normalized hand image that can be joined to the shape parameters returned by the AAM fitting process. This join of vectors can be used as a unique vector which contains aligned features. Therefore, by using a small quantity of images, it is possible to reach acceptable classification rates as we have demonstrated. As a future work, we propose to investigate approaches for reducing the quantity of examples required for incremental learning. Redundant examples should be avoided in such a way that only novel information could enter to the system.

References

1. Greulich, W., Pyle, S.: Radiographic Atlas of Skeletal Development of Hand and Wrist, 2nd edn. Stanford University Press, Stanford (1971)
2. Tanner, J., Whitehouse, R., Cameron, N., Marshall, W., Healy, M., Goldstein, H.: Maturity and Prediction of Adult Height (TW2 Method), 2nd edn. Academic Press, London (1975)
3. Hsieh, C.W., Jong, T.L., Chou, Y.H., Tiu, C.M.: Computerized geometric features of carpal bone for bone age estimation. Chin. Med. J. **120**, 767–770 (2007)
4. Molinari, L., Gasser, T., Largo, R.H.: Tw3 bone age: Rus/cb and gender differences of percentiles for score and score increments. Ann. Hum. Biol. **31**, 421–435 (2004)
5. Niemeijer, M., van Ginneken, B., Maas, C., Beek, F., Viergever, M.: Assessing the skeletal age from a hand radiograph: automating the tanner-whitehouse method. In: Sonka, M., Fitzpatrick, J. (eds.) SPIE Medical Imaging, vol. 5032, pp. 1197–1205. SPIE (2003)
6. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Comput. Vis. Image Underst. **61**, 38–59 (1995)
7. Aja-Fernndez, S., de Luis-Garcia, R., Martin-Fernandez, M.A., Alberola-Lpez, C.: A computational TW3 classifier for skeletal maturity assessment. A computing with words approach. J. Biomed. Inf. **37**, 99–107 (2004)
8. Liu, H., Chou, Y., Tiu, C., Lin, C., Chen, C., Hwang, C., Hsieh, C., Jong, T.: Bone age pre-estimation using partial least squares regression analysis with a priori knowledge. In: 2014 IEEE International Symposium on Medical Measurements and Applications, MeMeA 2014, Lisboa, Portugal, pp. 164–167, 11–12 June 2014
9. Adeshina, S.A., Cootes, T.F., Adams, J.E.: Evaluating different structures for predicting skeletal maturity using statistical appearance models. In: Proceedings of MIUA (2009)
10. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 681–685 (2001)
11. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
12. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**, 71–86 (1991)
13. Luo, B., Hancock, E.: Iterative procrustes alignment with the EM algorithm. Image Vis. Comput. **20**, 377–396 (2002)
14. Ross, A.: Procrustes analysis. Technical report, Department of Computer Science and Engineering, University of South Carolina, SC 29208 (2004). www.cse.sc.edu/songwang/CourseProj/proj2004/ross/ross.pdf
15. Passerini, A.: K-nearest neighbour learning. Department of Information Engineering and Computer Science. University of Trento, Italy (2015). http://disi.unitn.it/passerini/teaching/2015-2016/MachineLearning/slides/02b_nearest_neighbour/talk.pdf
16. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. **46**, 175–185 (1992)