ISSN:0975-9646

R.Srinivas et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (5) , 2013, 692-695

Digital Envelope for MRI Scanning by using B-Spline and Gaussian Densities

R.Srinivas, M. Rajesh Kumar

AITAM College of Engineering,
SRIKAKULAM (A.P).

Abstract- In this paper we present a method for unsupervised image clustering and B-spline curve interpolation for remove the noise. Images are clustered such that the mutual information between the clusters and the image content is maximally preserved. The clustering principle is applied to both discrete and continuous image representations, using Gaussian densities. We find that whole-brain analysis in this manner allows automatic classification of images based on Disease if the whole brain is included, If a sequence of planar points and vectors are given then a free-form curve can calculated which interpolates the points and has the given tangent vectors in these points. Our method gives a fast interpolation of these data using extra control points. Then we provide a method which allows to interpolate the same set of data without any predefined order of the points, i.e. a set of scattered points with the vectors.

Keywords: Unsupervised image clustering, B-spline curve interpolation, mixture of Gaussians.

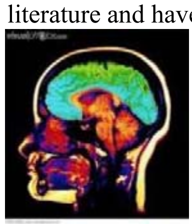
INTRODUCTION

Image clustering and categorization is a means for high-level description of image content. The goal is to find a mapping of the archive images into classes (clusters) such that the set of classes provide essentially the same prediction, or information, about the image archive as the entire image set collection. The generated classes provide a concise summarization and visualization of the image content. Image archive clustering is important for efficient handling (search and retrieval) of large image databases. In the retrieval process, the query image is initially compared with all the cluster centers. The subset of clusters that have the largest similarity to the query image is chosen, following which the query image is compared with all the images within this subset of clusters. Search efficiency is improved due to the fact that the query image is not compared exhaustively to all the images in the database. Image clustering may be performed using discrete image representations (e.g. histograms) as well as continuous image representations (e.g. probabilistic continuous image modeling based on mixture of Gaussian densities). In recent work, that compares between various image representation schemes, image modeling based on mixture of Gaussian densities was shown to outperform discrete image representations (such as the well-known color histograms, color correlograms, and more) . In the current work we demonstrate unsupervised clustering in both the discrete and continuous image representations domains. The clustering method presented in this work is based on the information bottleneck (IB) principle. Characteristics of the proposed method include: 1) Image *models* are clustered rather than raw image pixels (image models may be discrete Or continuous). 2) The IB method provides a simultaneous construction of both the clusters and the distance measure between them. 3) A natural termination of the bottomup clustering process can be determined as part of the IB principle. This provides an automated means for finding the relevant number of clusters per archive. 4) The continuous agglomerative version of the IB clustering scheme is extended to include relaxation steps for better clustering results.

Rational B-spline curves and surfaces as the generalization of B-spline curve and surface are widely used in image processing system. Basically these methods have been developed for approximating points, but they can be used as interpolating curves or surfaces as well. In this paper we will use the rational B-spline curve for a special interpolation problem, where beside the points the tangent vectors of the future curve are also given. The method is similar to the case of B-spline: the control points of the future curve is calculated from the given data, so finally it will be an approximating curve, but given points will be on the curve and it will have the given tangent vectors. This problem can also be formulated without giving the order of points. Since all the basic free-form methods are defined with a sequence of points as input data, in this case we use an artificial neural network to order the points and then we apply the method mentioned above.

Grouping pixels into clusters and remove the noise using Bezier curves

In the first layer of the grouping process the raw pixel representation of an input image is shifted to a mid-level representation. The image representation may be discrete (e.g. histograms) or continuous. Histograms are well known in the literature and have been used substantially .



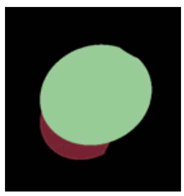


Figure 1: Input image (left) Image modeling via Gaussian mixture (right)

In this section we briefly introduce the more recently proposed continuous image representation schemes. In the continuous domain, each image is modeled as a mixture of Gaussians in the color feature space. In order to include spatial information, the $(x; y)$ position of the pixel is appended to the feature vector. Following the feature extraction stage, each pixel is represented with a five-dimensional feature vector, and the image as a whole is represented by a collection of feature vectors in the five dimensional space. Pixels are grouped into homogeneous regions, by grouping the feature vectors in the selected feature space. The underlying assumption is that the image colors and their spatial distribution in the image plane are generated by a mixture of Gaussians. Each homogeneous region in the image plane is thus represented by a Gaussian distribution, and the set of regions in the image is represented by a Gaussian mixture model. The distribution of a d-dimensional random variable is a mixture of k Gaussians if its density function is:

$$f(y) = \sum_{j=1}^{k} \alpha j \frac{1}{\sqrt{(2\pi)^{d} |\Sigma j|}} \exp\{-\frac{1}{2}(y-\mu_j)^T \Sigma_j^{-1}(y-\mu_j)\}$$

The Expectation-Maximization (EM) algorithm is used to determine the maximum likelihood parameters of a mixture of k Gaussians in the feature space (similar to). The Minimum Description Length (MDL) principle serves to select among values of k. In our experiments, k ranges from 4 to 8. Figure 1 shows two examples of learning a GMM model for an input image. In this visualization each localized Gaussian mixture is shown as a set of ellipsoids. Each ellipsoid represents the support, mean color and spatial layout, of a particular Gaussian in the image plane.

ALGORITHM

The algorithm starts with the trivial clustering where each cluster consists of a single point. In order to minimize the overall information loss caused by the clustering, classes are merged in every (greedy) step such that the loss in the mutual information caused by merging them is the smallest. Let $c1$ and $c2$ be two clusters of symbols from the alphabet of X, the information loss due to the merging of $c1$ and $c2$ is:

$$d(c1; c2) = I(C_{before}; Y) \text{ ¡ } I(C_{after}; Y) \geq 0$$

where $I(C_{before}; Y)$ and $I(C_{after}; Y)$ are the mutual information between the classes and the feature space before and after $c1$ and $c2$ are merged into a single class. Standard information theory manipulation reveals:

$$d(c1; c2) = \sum_{y,i=1,2} P(c_i,y) \log\frac{P(c_i,y)}{P(c_i)P(y)} - \sum_{y} P(c_1Uc_2,y) \log\frac{P(c_1Uc_2,y)}{P(c_1Uc_2)P(y)}$$

$$= \sum_{y,i=1,2} P(c_i,y) \log\frac{P(y|c_i)}{P(y|c_iUc2)}$$

$$= \sum_{y,i=1,2} P(c_i)\, D_{KL}\, P(y|c_i) || P(y|c_iUc2)$$

Hence, the distance measure between clusters $c1$ and $c2$, derived from the IB principle, takes into account both the dissimilarity between the distribution $p(y_j,c_1)$ and $p(y_j,c_2)$ and the size of the two clusters. The greedy AIB algorithm arranges the objects in a tree structure, which has many advantages for database management.

The algorithm also enables to define the optimal number of clusters that represent the objects in the database. However, the main obstacle to the greedy agglomerative procedure is that finding an optimal clustering solution is not guaranteed. In fact, it is not guaranteed to find a stable solution, in which each object belongs to the cluster it is most similar to. The issue of cluster optimization is common in many (both top-down and bottom-up) hierarchical clustering techniques. These techniques, due to their greedy nature, often require additional relaxation steps for cluster optimization.

INTERPOLATION OF SCATTERED POINTS WITH TANGENT VECTORS

This problem is similar to the previous one, but the given points have no predefined order, i.e. we do not know which point has to be the first and which one is the last one. Since the rational B-spline method can be applied only on a sequence of points (and weights), first of all we have to order the points. For this purpose an artificial neural network will be used. After this step the same procedure described above can be applied. Now after a short description of the applied net, the Kohonen network the ordering process and the interpolation will be discussed. For more detailed discussion of the ordering method by Kohonen network see. The Kohonen neural network is a two-layered non-supervised learning neural network. Self organizing networks, like the applied Kohonen net, organize the input data during the so called learning phase without any supervision. The most important part of the algorithm is the training rule, which modifies the network according of the input points. Let a set of points Pi(i = 1, . . . , n) (scattered data) and a set of vectors ti(i = 1, . . . , n) be given on the plane. Our first task is to determine the order of the points for the interpolation problem.

Miklos Hoffmann and Emőd Kovacs the Kohonen net is used to order the points. The first layer of neurons is called input layer and contains the two input neurons which pick up the data, the planar points. The input neurons are entirely interconnected to a second, competitive layer, containing m neurons (where m _ n, usually m = 4n). The weights associated with the connections are adjusted during training by the following rule: — Coordinates of the scattered points: Pi(x1i, x2i, x3i)

(i = 1, . . . , n) — Coordinates of the output points: Qj (w1j ,w2j ,w3j)

(j = 1, . . . ,m)

STEP 1. Initialize the weights wsj , (s = 1, 2, 3 j = 1, . . . ,m) as small random values around the average of the coordinates of the input points. Let the training time t = 1

STEP 2. Present new input values (x1i0 , x2i0 , x3i0), as the coordinates of a randomly selected input point pi0

STEP 3. Compute the Euclidean distance of all output nodes to the input point:

dj $=\sum_{s=1}^{3}(x_{si0} - w_{sj})^2$

STEP 4. Find the winning unit qj0 as the node which has the minimum distance to the input point, so where j0 is the value for which

dj0 = min (dj)

STEP 5. Compute the neighborhood N(t) = (j0, j1, . . . , jk)

STEP 6. Update the weights (i.e. the coordinates) of the nodes in the neighborhood by the following equation:

Wsj (t + 1) = wsj (t) + $\eta(t)$(xsi0 − wsj (t)) ∀j ∈ N(t)

Where $\eta(t)$ is a so called gain term, a Gaussian function decreasing in time.

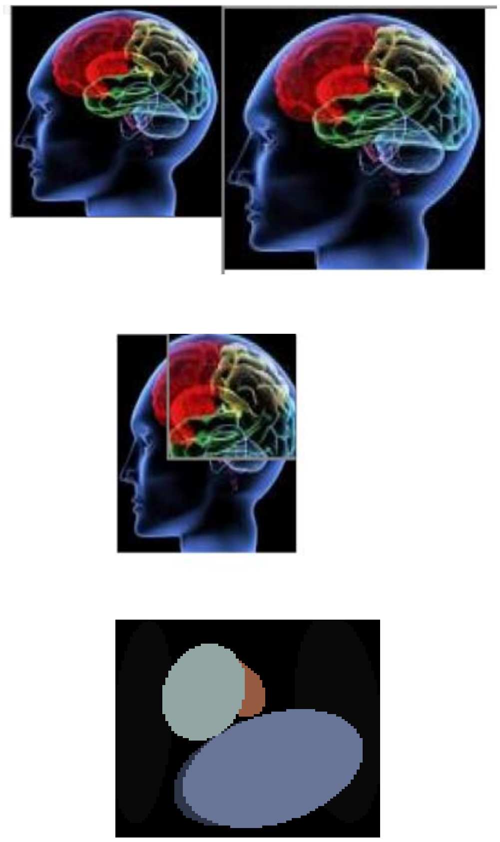
STEP 7. Let t = t + 1. Repeat STEP 2-7 until the network is trained.

The network is said to be trained if all the input points are on the polygon, that is for all the input points Pi(i = 1, . . . ,m) there is an output vector oj such that after a certain time t0 the Euclidean distance of oj and Pi is smaller than a predefined limit. A stronger convergence can be obtained if we require that the output vectors which do not converge to an input vector be on the line determined by its two neighbouring output vectors. This stronger convergence is important especially in term of the smoothness of the future curve. After the ordering process the same algorithm can be applied to calculate the interpolation curve as we described above. At this part of the process it is irrelevant, that the input points were scattered.

RESULTS

This section presents an investigative analysis of the IB method for image clustering. Experimental results demonstrate the IB method's ability to discover an optimal number of clusters in the database. Retrieval experiments are used to evaluate the clustering quality of the proposed method and of the various clustering algorithms introduced. The database used throughout the experiments consists of several images selectively hand-picked from the MIT database to create different categories.

The images within each category have similar colors and color spatial layout, and can be labeled with a high-level semantic description. The bottom-up clustering method was applied to our database of several images. The clustering is performed on the GMM image representation. We started with several clusters where each image model is a cluster. After several steps all the images were grouped into a single cluster. The given database was thus arranged in a tree structure. The loss of mutual information during each merging step of the clustering process is shown in Figure (i, ii). After zoom the reports experts can easily identify the diseases occurred in report we are perform two different types of zooming those are basic zoom and lens zoom. The labels associated with the image indicate the number of clusters created in the corresponding step. There is no need to present the information loss during the entire clustering process, since meaningful changes occur only towards the end of the process. There is a gradual increase in information loss until we reach a point of significant loss of information. This point helps us determine a meaningful" number of clusters existing in the database. From this point on, every merge causes a significant degradation of information and therefore leads to a worse clustering scenario. The first significant jump in the graph is found in the transition from different clusters.



(i)Basic Zoom (ii) lens Zoom (iii) ellipse

CONCLUSION

In this paper, we have tried to make use of the nature of B-splines $N_{ik}(t)$ to improve the performance of the existing parameterizations. First, parameter values are used as knot values. It gives us nice-looking curves. More specifically, small bulges are obtained between data points. The computation of B-spline interpolation is faster and simpler. However, this scheme doesn't work well in cases of other than order 4.On the other hand, universal parameterizations gives us more natural looking curves while the curves are transformation invariant. The computation of interpolation is much simpler and faster because the selection of parameter and knot values does not depend on data points at hand.

Image variations including illumination irregularities, texture and other artifacts are not accounted for in the models used. The additional features influence on clustering quality should be investigated. Future work entails making the current method more feasible for large databases and using the tree structure created by the AIB algorithm, for the Creation of a "user friendly" browsing environment.

REFERENCES

[1] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley, New York, 1991.
[2] Nielson, G.M. (1987), Coordinate free scattered data interpolation, in: Chui, C., Schumaker, L.L. and Utreras, F., eds., *Topics in Multivariate Approximation*, Academic Press, 175–184.
[3] Hoschek, J. and Lasser, D. (1993), *Fundamentals of Computer Aided Geometric Design*, A K Peters.
[4] J. Chen, C.A. Bouman, and J.C. Dalton. Hierarchical browsing and search of large image databases. *IEEE transactions on Image Processing*, 9(3):442–455, March 2000.
[5] L. Piegl, W. Tiller: The NURBS Book, Springer Verlag, 1997.
[6] Raghu K, Keller JM: A possibilistic approach to clustering. IEEE Trans Fuzzy Syst 1993, 1(2):98-110.
[7] Mignotte M: Segmentation by fusion of histogram-based K-means clusters in different color spaces. IEEE Trans Image Process 2008, 17(5):780-787.
[8] Dempster AP: Upper and lower probabilities induced by multivalued mapping. Ann Math Stat 1967, 38:325-339.
[9] Shafer G: A Mathematical Theory of Evidence Princeton University Press, Princeton, NJ; 1976.