Implementation of Radial Basis Function Network in Content-Based Video Retrieval

Purushothaman Srinivasan
Adama Science and Technology University

Implementation of Radial Basis Function Network in Content-Based Video Retrieval

S. Prasanna, S. Purushothaman and R. Rajeswari

Abstract This paper presents retrieving a video from a given database using radial basis function (RBF) network method. The features of the videos are used by RBF for training and testing RBF in the algorithm developed. The features of frames of a video are extracted from the contents in the form of text, audio, and image. In this analysis, RBF is programmed to retrieve the words spoken by four different speakers in video presentation.

Keywords Content-based retrieval · Radial basis function · Video retrieval

Introduction

Video will be one of the key issues in the upcoming Internet evolution in infotainment and education. Converting raw video streams into highly and thoroughly structured and indexed, web ready, database-driven information entities are a must. Information databases have evolved from simple text to multimedia with video, audio, and text. The query mechanism for such a video database is similar in concept compared to a textual database. Object searching is analogous to word searching; scene browsing is similar to paragraph searching; video indexing is comparable to text indexing or bookmarking. Implementation for video

S. Prasanna (✉)
Department of MCA, VELS University, Chennai 600117, India
e-mail: prasannabalalak05@gmail.com

S. Purushothaman
PET Engineering College, Tirunelveli, Tamil Nadu 627117, India
e-mail: drsppuru@gmail.com

R. Rajeswari
Research Scholar Mother Teresa Women's University Kodaikanal, Kodaikanal, India
e-mail: rajeswaripuru@gmail.com

S. Sathiakumar et al. (eds.), *Proceedings of International Conference on Internet Computing and Information Communications*, Advances in Intelligent Systems and Computing 216,
DOI: 10.1007/978-81-322-1299-7_42,

content-based searching is very different and much more difficult than the query mechanism for a textual database. Scene change detection technique can often be applied for scene browsing and automatic and intelligent video indexing of video sequences for video databases. Once individual video scenes are identified, we use content-based indexing mechanisms (such as indexing by object texture, shape, color, motion, etc.) to index and query image contents in each video scene.

Related Works

There are many information carriers in a video stream, as is the visual content, the narrative or speech part, text captions. Visual content remains the most important. We prefer representative image from a long scene with little -or no- change. This process is the key frame extraction. Two kinds of key frame extraction strategies have been developed and used by various researchers. The simplest way is to select one or several frames from each segmented shot. Some use the first frame of each shot as the representative frame, i.e., the key frame of the shot. Others may use a random one, the last one, or the middle one as the prototypical frame.

Lots of research works have been done by earlier researchers on retrieval of videos. Web browsers and e-mail service providers have developed many indexing methods and retrieval algorithms for retrieving videos. In addition to the existing development, we are focusing on implementing intelligent method like radial basis function (RBF).

Erol and Kossentini [1] use object-based video representation, such as the one suggested by the MPEG-4 standard, offers a framework that is better suited for object-based video indexing and retrieval. In such a framework, the concept of a “key frame” is replaced by that of a “key video object plane.” Approaches mainly differ in the set of acoustic features used to represent the audio signal and the classification technique applied [2]. In speech research, features such as Mel frequency cepstral coefficients (MFCCs) or linear prediction coefficients (LPCs) have been demonstrated to provide good representations of a speech signal allowing for better discrimination than temporal- or frequency-based features alone. Holmes and Holmes [3] and Liu and Wan [4] conclude from their study that cepstral features, such as MFCCs perform better than temporal- or frequency-based features and advocate their use for general audio tasks particularly when the number of audio classes is large.

Semantic filtering and retrieval of multimedia content is crucial for efficient use of the multimedia data repositories. Naphade and Huang [5] and Slaney [6] proposed a state-of-the-art system which incorporates a mapping between audio and semantic spaces. Methods are developed to describe general audio with words (and also predict sounds given a text query) using a labeled sound set. Adams et al. [7] have focused on semantic classification through the identification of meaningful intermediate-level semantic components using both audio and video features.

Fang et al. [8] proposed indexing and retrieval system of the visual contents based on feature extracted from the compressed domain. Direct possessing of the compressed domain spares the decoding time, which is extremely important when indexing large number of multimedia archives. Benmokhtar and Huet [9] proposed an improved version of RBF network based on Evidence Theory (NNET) using one input layer and two hidden layers and one output layer, to improve classifier combination and recognition reliability in particular for automatic semantic-based video content indexing and retrieval. Many combination schemes have been proposed in the literature according to the type of information provided by each classifier as well as their training and adaptation abilities.

Snoek et al. [10] identified three strategies to select a relevant detector from thesaurus, namely: text matching, ontology querying, and semantic visual querying for a given query. They evaluate the methods against the automatic search task of the TRECVID 2005 video retrieval benchmark, using news video archive of 85 h in combination with a thesaurus of 363 machine learned concept detectors. They assessed the influence of thesaurus size on video search performance, evaluated and compared the multimodal selection strategies for concept detectors, and finally discuss their combined potential using oracle fusion.

Lu et al. [11] surveyed some of the existing techniques and systems for content-based indexing and retrieval of two main types of multimedia data images and videos. In content-based image retrieval, they have proposed multiscale color histograms by incorporating color and spatial information.

Gao et al. [12] discussed video sequences as temporal trajectories via scaling and lower dimensional representation of the video frame luminance field, and a video trajectory indexing and matching scheme was developed to support video clip search. Simulation results demonstrated that the proposed approach achieved excellent performance in both response speed and precision-recall accuracy.

Slaney [6] proposed a state-of-the-art system which incorporates a mapping between audio and semantic spaces. Methods are developed to describe general audio with words (and also predict sounds given a text query) using a labeled sound set. Adams et al. have focused on semantic classification through the identification of meaningful intermediate-level semantic components using both audio and video features.

Methods and Materials

Material

Table 1 shows the number of speeches recorded at two different instances

In this work, the audio track of videos is extracted from each shot and stored as patterns. When an user gives audio as input query, it will match with the stored patterns using RBF.

Table 1 Recordings of speech at different instances

Person name	Speech1	Speech2
Prasanna	√	√
Purushothaman	√	
Rajeswari	√	
Shwetha	√	

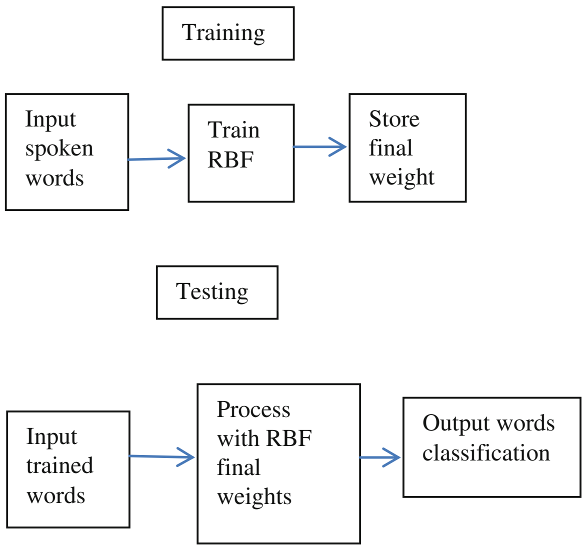

Fig. 1 Training and testing of radial basis function (RBF)

Schematic Diagram

Figure 1

Radial Basis Function Network

Figure 1 represents the block diagram of radial basis function (RBF) algorithm. A RBF is a real-valued function whose value depends only on the distance from the origin. If a function 'h' satisfies the property $h(x) = h(\|x\|)$, then it is a radial function. Figure 2 depicts the training method used through neural networks algorithm using a single hidden layer method. Tables 2 and 3 provide the training and testing procedure employed in doing so. Their characteristic feature is that their response decreases (or increases) monotonically with distance from a central point. The center, the distance scale, and the precise shape of the radial function are parameters of the model, all fixed if it is linear. Here the distance is found between a pattern and each center. The center is also one of the patterns

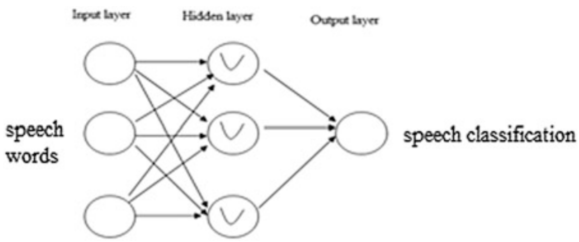

Fig. 2 Training with RBF

Table 2 Training RBF

Step 1: radial basis function (RBF) is applied
No. of input = 4
No. of patterns = 10
No. of center = 10
Calculated RBF as
RBF = exp (-X)
Calculated Matrix as G = RBF
$A = G^T * G$
Calculated $B = A^{-1}$
Calculated $E = B * G^T$
Step 2: Calculated the Final Weight F = E * D
Step 3: Stored the Final Weight in a file.

Table 3 Testing RBF

Step 1: Read the input
Step 2: Read the final weights
Calculated Numerals = F * E
Step 3: Input to RBF
Step 4: Classification of Speech by RBF

predefined. The square of the distance is a node in the hidden layer. An exponential function is used as an activation function which will be the output of the particular node. The number of nodes in the hidden layer is based on the number of centers decided in an implementation.

Results and Discussion

Figure 3 represents the speech patterns versus the prediction of words by the various speakers.

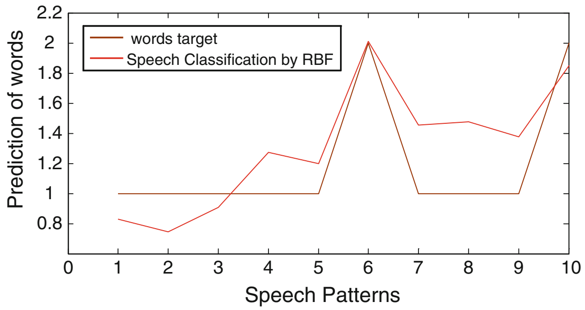

Fig. 3 Speech patterns vs prediction of words

Conclusion

In this paper, we discuss the implementation of RBF and the distance between pattern and centers are found. Final weights are calculated in Tables 2 and 3. The RBF network learns the patterns through iteration.

References

1. Erol, B., Kossentini, F.: Automatic key video object plane selection using the shape information in the Mpeg-4 compressed domain. IEEE Trans. Multimedia **2**(2), 129–138 (2000)
2. Tsekeridou, S., Pitas, I.: Content-based video parsing and indexing based on audio–visual interaction. IEEE Trans. Circuits Syst. Video Technol. **11**(4), 522–535 (2001)
3. Holmes, J.N., Holmes, W.J.: Speech Synthesis and Recognition, 2nd ed. Taylor & Francis, London (2001)
4. Liu M., Wan C.: A study on content-based classification and retrieval of audio database. In: Proceedings of International Database Engineering and Applications Symposium, Grenoble, France. IEEE Comput. Soc., Washington DC pp. 339–345 (2001)
5. Naphade, M.R., Huang, T.S.: A probabilistic framework for semantic video indexing, filtering, and retrieval. IEEE Trans. Multimedia **3**(1), 141–151 (2001)
6. Slaney M.: Mixtures of probability experts for audio retrieval and indexing. In: Proceedings of IEEE International Conference on Multimedia and Expo, Goteborg, Sweden, vol. 1, pp. 345–348 (2002)
7. Adams, W.H., Iyengar, G., Lin, C.Y., Naphade M.R., Neti, C., Nock, H.J., Smith J.R.: Semantic indexing of multimedia using audio, text and visual cues, EURASIP J. Appl. Signal Process. **2**, 170–185 (2003)
8. Fang, H., Qahwaji, R., Jiang, J.: Video indexing and retrieval in compressed domain using fuzzy-categorization, ISCV 2006, LNCS 4292, Springer-Verlag Berlin Heidelberg, pp. 1143–150 (2006)
9. Benmokhtar, R., Huet, B.: Neural network combining classifier based on Dempster-Shafer theory. In: Proceedings of the International Multiconference on Computer Science and Information Technology, Singapore, pp. 3–10 (2006)
10. Snoek, C.G.M., Huurnink, B., Hollink, L., De Rijke, M., Schreiber, G., Worring, M.: Adding semantics to detectors for video retrieval. IEEE Trans. Multimedia **9**(5), 975–986 (2007)

11. Lu, H., Xue, X., Tan, Y.-P.: Content-Based Image and Video Indexing and Retrieval, pp. 118–129. Springer Verlag, Berlin Heidelberg (2007)
12. Gao, L., Li, Z., Katsaggelos, A.: An efficient video indexing and retrieval algorithm using the luminance field trajectory modeling. IEEE Trans. Circuits Syst. Video Technol., **19**(10), 1566–1570 (2009)