ISSN 2047-3338

An Effective Image Retrieval System Using Textual and Visual Properties

S. Swetha and Prof. K. Ashok Babu

***Abstract*— Image retrieval is a challenging task that requires efforts from image processing, link structure analysis, and web text retrieval. Since content-based image retrieval is still considered very difficult, most current large scale web image search engines exploit text and link structure to understand the content of the web images. However, local textual information, such as caption, filenames and adjacent text, is not always reliable and informative. And also, there is no commercial web image search engine support RF because of scalability, efficiency and effectiveness. Therefore, global texture information should be taken into account to support RF and a web image retrieval system makes relevance judgment. In this paper, we propose a re-ranking method to improve web image retrieval by reordering the images retrieved from an image search engine using RF. The re-ranking process should be applicable to any image search engines with little effort and experimental results on a database contain three million web images to show RF is effective.**

***Index Terms*— Image Retrieval, Relevance Feedback, Implicit Feedback and RF Fusion**

I. INTRODUCTION

BY the growth of both worldwide and the number of digital images there is an urgent need for image retrieval systems. The most popular commercial search engines, such as Google, Yahoo and Alta Vista support image retrieval by keywords e.g., pic search. A common limitation for the most of the commercial search engines are their search process is passive, there is no interaction between user and system. In order to make system active, the system could take advantage of relevance feedback techniques.

The main criteria of the relevance feedback are whenever the thing is passive (which does not respond to the given system). It provides the activeness (which will responds to the given system). Common limitation for the most of the search engines are their search process is passive in order to avoid this drawback we are going for the relevance feedback which makes the system active.

Relevance feedback is an online learning technique, originally developed for information retrieval [3] system to improve the effectiveness of the information retrieval system.

The main aim of relevance feedback is to provide the interaction between the user and system. During retrieval process, the user interacts with the system and rates the relevance of the retrieved documents, according to user judgment. With this additional information, the system dynamically learns the users' intention, and gradually presents better results. Since the introduction image retrieval during mid -1990's, it has attracted tremendous attention in the CBIR community and has been shown to provide dramatic performance improvement [10]. However, almost all the existing relevance feedback algorithms in image retrieval systems are performed in an explicit way. It is noted that explicit relevance feedback techniques have been underutilized as they place an increased cognitive burden on users while the benefits are not always obvious to them. Comparing with explicit feedback, implicit feedback could be collected at much lower cost, in much larger quantities and without burden on the user. As one of the most effective implicit feedback information, click-through data has been used either as absolute relevance judgments [8]. During this RF process, it provides the better experimental results.

In this paper, we propose a relevance feedback technique to retrieve the web images. This RF performs three steps to retrieve the relevant images.

1) During first step, we combine textual feature based RF (TBRF) and visual feature based RF (VBRF)in a sequential way using RF fusion. In this a TBRF is first performed to select a possibly relevance image set. Then, VBRF is combined with TBRF to further re-rank the resulting web images. The fusion of VBRF and TBRF is query concept dependent and automatically selects the relevant images.

2) During second step, the textual-feature based RF is performed using metadata. Then we could integrate RF into web image retrieval in a practical way.

3) During third step, a new graphical user interface is performed to support implicit RF. The GUI provides the users click-through data related to the implicit relevance feedback in order to release the burden from users.

The textual features of the commercial search engines are represented as metadata which includes filename, category, ALT text, URL and surrounding text of the images. This metadata is used as the textual source for the textual space construction. To build the textual space two approaches are in

S. Swetha, pursuing M.Tech in Digital Electronics and Communication Systems from Sri Indu College of Engineering and Technology, India

Dr. K. Ashok Babu, Professor and Head of the Department, with the Sri Indu College of Engineering and Technology, India

work. One is the straight forward approach, directly using the above metadata to obtain the textual feature .But straightly using the textual information leads to a time consuming and suffers from noisy terms. Another one is based on Search result clustering [1] algorithm is used to construct textual space. In this paper, RF uses the second approach to avoid heavy computations over millions of retrieved images provides an efficient and effective mechanism to construct an accurate and low dimensional textual space.

The commercial web image retrieval systems solely depend on the textual information. Web images are characterized by both textual and visual features. The textual

Representation of an image is insufficient compared to the visual content present of the image. The visual features provides finer granularity of image. Considering the characteristics of both textual and visual, RF in textual could guarantee the relevance and RF in visual could meet the need for finer granularity. Thus, in this paper we propose RF for image retrieval using both textual and visual features in a sequential way.

In this paper, we employ implicit feedback to overcome the limitation of explicit feedback techniques where an increased cognitive burden is placed on the users. In this, implicit feedback using click- through data has not been applied to web image retrieval systems. Comparing with text retrieval, image retrieval has the following two characteristics. First, the thumbnail of an image reflects more information than the title and snippet of a Web page, so click-through information of image retrieval tends to be less noisy than that of text retrieval. Second, unlike textual document, the content of an image can be taken in at a glance. As a result, the user will possibly click more results in image retrieval than in text retrieval. Both characteristics imply that click-through data could be helpful for image retrieval.

The remainder of this paper is organized as follows. In section 2, we describe the image retrieval using RF mechanism. In section 3, we describe the User interface. Experimental results are presented and analyzed in section 4. Finally we conclude and discuss future work in section 5.

II. Image Retrieval using RF

A. Image Representation

The images are collected from several photo forum sites, e.g., photo sig. Every image contains rich metadata like title, category, photographer's comment and other people's critiques. This is the dataset for Relevance feedback mechanism. For example a photo of photo sig has the following metadata.

- Title: early morning
- Category: landscape, nature, rural
- Comment: I found this special light one early morning in pyreness along the Vicdessos River near our house…
- One of the critiques: wow, I like this picture very much; the light is great on the snow and on the sky

To textually represent an image, vector space model [6] with TFIDF weighting scheme is adopted. In this, the textual feature of an image I is an L-dimensional vector and it is denoted as,

$$\overrightarrow{F^T} = (w_1 \ldots\ldots w_L) \qquad (1)$$

$$\mathrm{W_i} = \mathrm{t}f_i . \log(N/n_i) \qquad (2)$$

Where:

$\overrightarrow{F^T}$ is the textual feature of an image.

Wi is the weight of the I th textual space.

L is the number of all distinct terms of all images in the textual space.

N is the total number of images.

$\mathrm{n_i}$ is the number of images whose metadata contains the term.

To represent an image in visual space, a 64 dimensional feature [4] was extracted. It is a combination of three features: (1) 6- dimensional color moments (2) 44 dimensional banded auto correlogram (3) 14 dimensional color texture moments. For color moments, the first two moments from each channel of CIE_LUV color space were extracted. For correlogram, the HSV color space with inhomogeneous quantization into 44 colors is adopted. The resulting image is a 64 dimensional vector,

$$\overrightarrow{F^V} = (f_1 \ldots\ldots f_{64}) \qquad (3)$$

And each feature dimension is normalized [0, 1] using Gaussian normalization.

B. Textual representation of RF

CBIR systems perform image retrieval based on the similarity defined in terms of visual features with more objectiveness. Although some new methods, such as the relevant feedback, have been developed to improve the performance of CBIR systems, low-level features do still play an important role and in some sense it will be the bottleneck for the development and application of CBIR techniques.

To represent RF in textual space, Rocchi's algorithm [3] is used. it is developed in the mid-1960's and has been proven to be one of the most effective RF algorithms in the information retrieval. The main aim of this algorithm is to construct an optimal query so that the difference between the average score of a relevant document and average score/ of a non-relevant document is maximized. Cosine similarity is used to calculate the similarity between an image and the optimal query. So only clicked images are available for our proposed retrieval process, we assume clicked images to be relevant and define the feature of optimal query as follows:

$$\overrightarrow{\vec{F}}_{\mathrm{opt}} = \vec{F}_{\mathrm{ini}} + \sum_{I \in Rel} \vec{F}_{\mathrm{I}} - \frac{\beta}{N_{Non-Rel}} \sum_{j \in Non-Rel} \vec{F}_{\mathrm{J}} \qquad (4)$$

Where:

$\vec{F}_{\mathrm{ini}}$ is the vector of the initial query.

$\vec{F_l}$ is the vector of the relevant image.

$\vec{F_j}$ is the vector of the non- relevant image.

Rel is the relevant image set.

Non-Rel is the non relevant image set.

N_{Rel} is the number of relevant images.

$N_{Non-rel}$ is the number of non relevant images.

α is the parameter for controlling the contribution of the relevant images and the initial query.

β is the parameter for controlling the contribution of non relevant images and the initial query.

In our case only relevant images are available for our proposed mechanism. So, we set α be 1 and β to be zero in our mechanism.

C. Visual representation of RF

To represent RF in visual space, visual properties of an image color, texture, shape are used. To develop this, Rui's [12] algorithm is used. Assume clicked images to be relevant, both an optimal query and feature weights are learned from the clicked images. And the feature vector of the optimal query is the mean of all features of clicked images. The weight of a feature dimension is proportional to the inverse of the standard deviation of the feature values of all clicked images [5]. Weighted Euclidean distance is used to calculate the distance between an image and the optimal query. Although Rui's algorithm is used currently, any RF algorithm using only relevant images could be used in the Image retrieval.

D. RF Fusion

Relevance feedback (RF) is a technique that helps searchers improve the quality of their query statements and has been shown to be effective in non-interactive experimental environments, and to a limited extent in IIR (Beaulieu, 1997). It allows searchers to mark documents as relevant to their needs and present this information to the IR system. The information can then be used to retrieve more documents like the relevant documents and rank documents similar to the relevant ones before other documents (Ruthven, 2001). RF is a cyclical process: a set of documents retrieved in response to an initial query are presented to the searcher, who indicates which documents are relevant. This information is used by the system to produce a modified query which is used to retrieve a new set of documents that are presented to the searcher. This process is known as an iteration of RF, and repeats until the required set of documents is found.

There has been some work on fusion of relevance feedback in different feature spaces [13]. A straightforward and widely used strategy is linear combination [2] [12]. Nonlinear combination using support vector machine (SVM) was proposed [13] in this paper, since the super-kernal fusion algorithm needs irrelevant images, it is incapable for systems only offering relevant images.

Considering, the textual features are more semantic-oriented and efficient than visual features while textual features have finer descriptive granularity than visual features. So, we combine the RF in both features in a sequential way. First, RF in textual space is performed to rank the initial resulting images using the optimal query. Then, RF in visual space is performed to re-rank the top images. The re-ranking process is based on a dynamic linear combination of the RF in both feature spaces.

The restricting of the re-ranking only on the top images has two advantages. First, the relevance of the top images could be guaranteed by the former RF in textual space. Second, the efficiency of RF process could be ensured. And RF in visual space could possibly be inefficient on a very large image set. The number of top images that affects both efficiency and effectiveness of the RF process is predetermined experimentally. The re-ranking process is based on a fusion of the RF in textual and visual spaces. The combination weights, that reflects the relative contribution of both spaces are automatically learned and query concept-dependent.

Assume there are n clicked imagesI_i, the similarity metric used to re-rank a top image I using RF in both textual and visual feature spaces is defined as follows:

$$S = \beta . S^V + (1- \beta)\ S^T \qquad (5)$$

$$\beta = \alpha\ .\exp(-\lambda\ . D_{avg}) \qquad (6)$$

$$D_{avg} = \Sigma_{i=0}^{n} \left\| \overrightarrow{F_i^V} - \overrightarrow{F_{opt}^V} \right\| / \text{ n} \qquad (7)$$

$$\overrightarrow{F_{opt}^V} = \Sigma_{i=0}^{n} \overrightarrow{F_i^V} / n \qquad (8)$$

$$S^V = 1- D^V \qquad (9)$$

Where:

S is the similarity metric in both textual and visual spaces.

S^V is the similarity between I's visual feature and $\overrightarrow{F_{opt}^V}$.

S^T is the cosine similarity between I's textual feature and $\overrightarrow{F_{opt}^V}$.

β is the dynamic linear combination parameter for similarity metric in both visual and textual spaces.

α and λ are the parameters which controls the Relative contribution of the RF in visual space.

D_{avg} is the deviation of the clicked image in the visual space.

$\overrightarrow{F_t^V}$ is the texture feature vector of the clicked image I_i.

$\overline{F_t^{opt}}$ is the feature vector of the optimal query in the visual space.

D^V is the weighted Euclidean distance between I's visual feature.

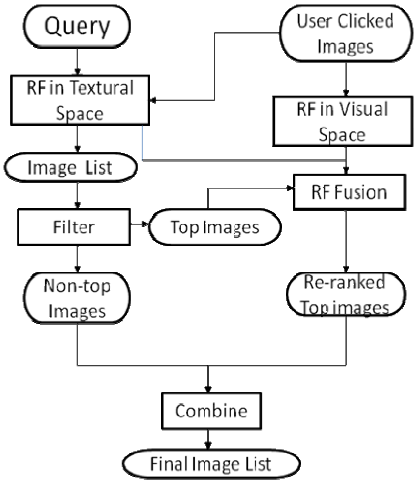

Fig. 1. Flow Chat of Relevance Feedback

From this, β tunes the visual feature's contribution to the overall similarity metric according to different query concept, α controls the overall contribution of RF in visual space, λ fine-tunes the contribution. If the query concept is well characterized by visual feature and the clicked images should be visually consistent, D_{avg} will be small (near 0). β should be large. Thus, visual feature will be important.

Since D_{avg} is query concept-dependent, the resulting combination parameter β is query concept-dependent as well. This property of the Parameter β results in a query concept-dependent fusion strategy for relevance feedback in both textual and visual space.

III. USER INTERFACE

To make the better use of the implicit feedback information, a new web image search UI named Mind Tracer is proposed. It consists of two types of pages: 1.main page and 2.detail page. The main page has three frames: search frame, recommendation frame, and result frame. The search frame contains an edit box for users to type query phrase. Only text-based queries are supported by Mind Tracer since they are friendly and familiar to the typical surfer of the web. After a user submits a query to Mind Tracer, the thumbnails of result images are shown in the result frame with five rows and four columns.

Initially, no images are shown in the recommendation frame. When the user clicks an image in the result frame for sake of user interest, the recommendation function is activated, so that the RF fusions are carried out.

As a result, a finer ranking of the initial results are obtained, and the top 20 recommended images will be shown in the recommendation frame. The images iteratively roll in the recommendation window with a scroll-bar that could be manually controlled by the user. And these are shown by the Fig. 2 and Fig. 3.

When the user's click through, the corresponding original image will be shown in a detail page. The detail page has two frames: image frame and snap shot frame.

If the user clicks another image in the result frame or in the recommendation frame, besides aforementioned system reactions, the recommended images will be shown in the snapshot frame of the detail page in case that the user wants more images from the former recommended image list. If the user clicks an image in the snapshot frame, the corresponding original image will be shown in the image frame. Once the user is satisfied with the recommended results, when the user click the refine button to move all the recommended images from recommendation frame to the result frame.

With the asynchronous scheme for refreshing the detail page and recommendation frame of the main page, no extra- waiting time is required for recommendation frame.

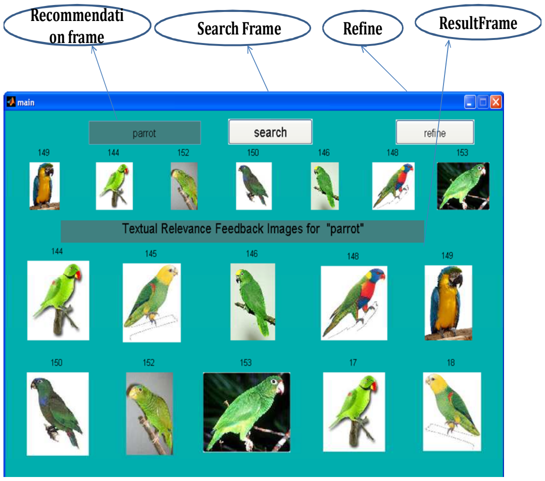

Fig. 2. Main Page of Mind Tracer

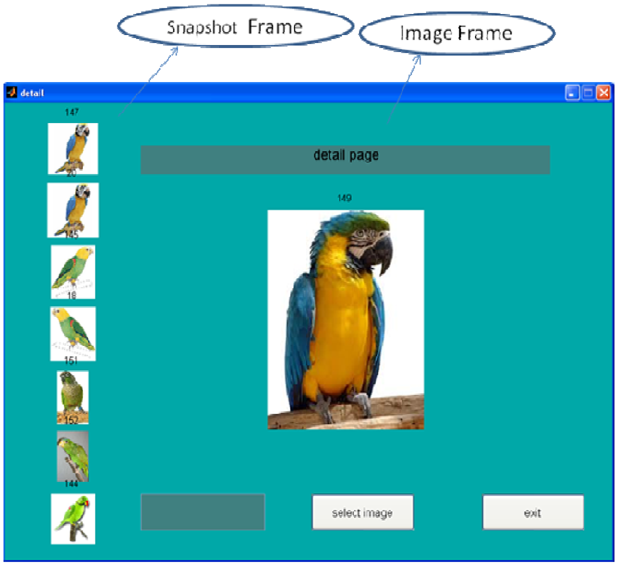

Fig. 3.Detail page of Mind Tracer

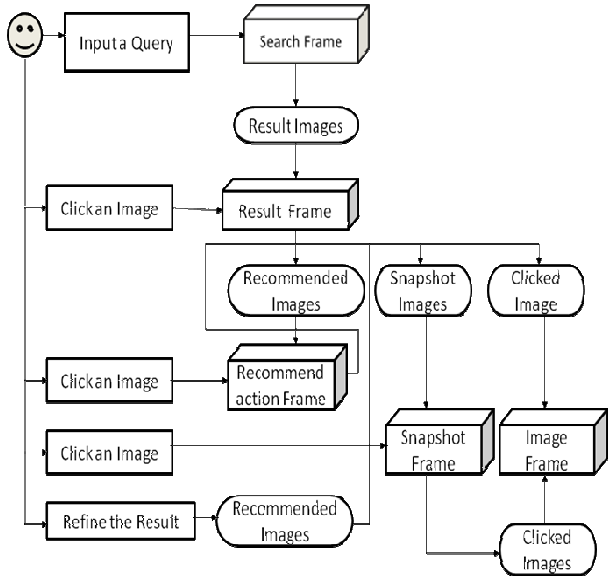

Fig. 4. Flow chart of User Interface

IV. EXPERIMENTAL RESULTS

A. Evaluation of Dataset

To construct the database, near three million images are crawled from several photo forum sites e.g., photosig. To automatically evaluate the RF algorithms, a subset of the web image was selected and manually labeled as follows.

First, Ten representative queries were chosen. Then, for each query, the key terms related to the top 20 images were identified. Finally, all resulting images of each query were manually annotated with the corresponding key terms.

From this, Ten representative queries include *Eagle(3809,17), Eiffel Tower(1517,24), Tiger(3826,22), Forest House(572,13) ,jaguar(337,13), Merry Christmas(5266,14,), Greek(2646,17), Rainbow(5376,17), Tulip(3743,13), pear(813,9).*

The numbers with in parentheses is the number of resulting images and the number of key terms for each query respectively. There are totally 159 key terms. For example, the key terms of query Tiger *include* butterfly, cat, cub, eye, flower, lily, people, Amoer, blue, common, dark, drinking, plain, Siberian, sitting, sleeping, small, Sumatran, swimming, white, yawning, young

In order to simulate the interaction between the user and a web image retrieval system, for each query Q_i, each related key term T_i was selected inturn to represent user's search intention. Images annotated with the term T_i were considered to be relevant to T_i. For each T_i, 5 iterations of user-and-system interaction were carried out. The system first ranked the initial resulting images using the optimal query learned in equation (2) and brought out the top k images using equation (3), the system examined the top 20 images image's to collect the relevant images which were regarded as click-through data. Those relevant images are labeled in previous iterations were directly placed in top ranks and excluded from the examining process. Precision is used as the basic evaluation measure. When the top 20 images are examined and there are N_{Rel} relevant images, the precision with in20 images is defined to be P(20) = N_{Rel} /20.

B. Evaluation of RF Fusion

The proposed RF fusion strategy (TVRF) has three parameters that need to be determined. In this, α controls the overall contribution of the RF in the visual space. λ fine-tunes the contribution and the scope K in which the resulting images are re-ranked by the combination of the textual and visual similarities. Because K is less correlated to α and λ .We first close K based on the simplified version by constraining λ to 0. i.e. S = $\alpha.S^V$+ (1-α)S^T. We conducted a series of experiments on varying the value of the α from 0-1. And K from 100 to 1000.

To verify the performance of the TVRF under the different values of α and K, K is finally set to be 200 which corresponds to the best result.

Then we fixed K to 200 and choose the α and λ simultaneously. We conducted another series of experiments by varying both α from 0 and 1. And λ from 1 to 256.To verify the detailed performance of the TVRF under different α and λ, we choose α=0.25 and λ =64 as the best parameters. To further validate k is fixed α to 0.25 and λ to 64.

C. Performance Comparison of RF

Four RF strategies were evaluated and compared: RF using textual feature only (TBRF), RF using visual feature only (VBRF), linear combination of RF in two features (LTVRF) and the proposed RF fusion strategy (TVRF). Figure 5 shows the F performance of four strategies for the ten representative queries and the average. The average precision of four strategies is 0.5481, 0.3905, 0.6705, and 0.883 respectively. From the result, it can be seen that TVRF performs the best among four strategies because it is capable of effectively combining textual and visual features.

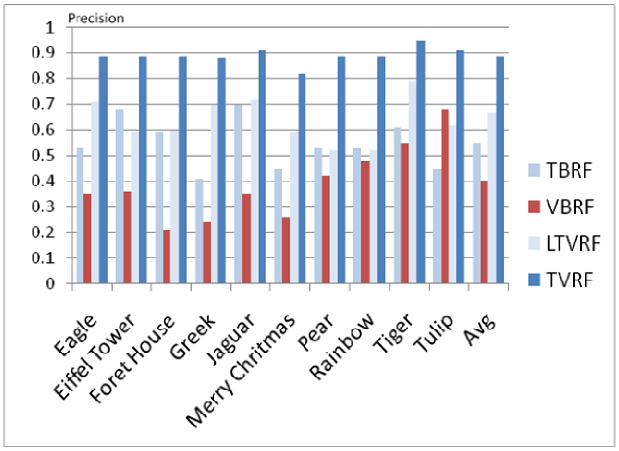

Fig . 5. Performances of four strategies

V. Conclusion

In this paper, we have presented effective image retrieval by using relevance feedback mechanism. During RF process, both textual and visual features are used in a sequential way. RF fusion strategy is proposed to combine RF in textual and visual space. To integrate RF into Image retrieval, the TBRF mechanism is employed. In this, implicit feedback, e.g., click through data has been proposed and a new graphical user interface is proposed to support implicit RF. By the experimental results, near three million web images show the effectiveness of the proposed mechanism.

REFERENCES

[1] E. Cheng et al., "Using Implicit Relevance Feedback to Advance Image Search," To appear in Proc. of ICME 2006.

[2] F. Jing et al., "A Unified Framework for Image Retrieval Using Keyword and Visual Features," IEEE Trans. on Image Processing, 14(7): 979-89, 2005.

[3] J. Rocchio, Relevance Feedback in Information Retrieval. Prentice-Hall, 1971.

[4] L. Zhang et al., "Efficient propagation for face annotation in family albums," Proc. of ACM Multimedia, pp.716-723, 2004.

[5] Q.K. Zhao et al., "Time-Dependent Semantic Similarity Measure of Queries Using Historical Click-Through Data," Proc. of WWW2006.

[6] R. Baeza -Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison-Wesley 1999.

[7] S. Sclaroff et al., "Image Rover: a content-based image browser for the World Wide Web," Proc. of IEEE Workshop on Content-based Access of Image and Video Libraries, pp.2-9, 1997.

[8] T. Joachim set al., "Accurately interpreting click through data as implicit feedback," Proc. of SIGIR, pp.154-161, 2005.

[9] T. Quack et al., "Cortina: a system for large-scale, content-based web image retrieval," Proc. of ACM Multimedia, pp.508-511, 2004.

[10] X.S. Zhou et al., "Relevance Feedback in Image Retrieval: A Comparative Study," ACM Multimedia Systems, 8(6):536-544, 2003.

[11] Y. Lu et al., "A Unified Framework for Semantics and Feature Based Relevance Feedback in Image Retrieval Systems," Proc. O f ACM Multimedia, pp. 31-38, 2000.

[12] Y. Rui et al. "Relevance feedback: A Power Tool for Interactive Content-based Image Retrieval," IEEE Trans. on CSVT, 13(4):811- 820, 1998.

[13] Y. Wu et al., "Optimal Multi modal Fusion for Multimedia Data Analysis," Proc. of ACM Multimedia, pp.572-579, 2004.

S. Swetha, graduated from Nirmal Engineering College, Mancherial in Electronics and Communication Engineering. Now pursuing Masters in Digital Electronics and Communication Systems from Sri Indu College of Engineering and Technology.

Dr. K. Ashok Babu, Professor and Head of the Department (ECE) and for his constant cooperation, support and for providing necessary facilities throughout the M. Tech program. He has 15 years of experience at B. Tech and M. Tech level and he is working as a Professor in Sri Indu College of Engineering and Technology.