acollectionofpresentations

Useforretrievingthe
presentationrecording

takeapictureof
apresentation

playbackthe
presentation
segment

Figure1.Indexingintopresentationswithadigita lcamerapicture.

2. MULTIMEDIACONFERENCEROOM SETUP

Our presentation room is equipped with an omni-directional audiovisual meeting recorder and a presentationrecorder.Themeetingrecordercapture s360-degreevideoat30framespersecondandviewselec tionis performed based on the sound directions [4]. The presentation recorder automatically captures what i s displayed on the presentation screen/projector with the timestamps. To support a wide range of resolutions , the VGAoutputofthepresenter'smachineisconnected to a scanconverter where the VGA output is converted to an NTSC signal, captured by a frame grabber and saved in JPEGformatat640 ×480resolution.Theconversionfrom digital to analog and analog to digital results in some quality degradation in the image but this allows us to guaranteethatthepresentationrecordercancaptur evideo key frames from any laptop. The output of the meet ing recorderandthepresentationrecorderaresynchron izedby time-stampswithpost-hocclock-skewcorrection.

3. IMAGEMATCHINGALGORITM

We utilize a number of image features for matching the contentsofadigitalcameraimagetothatofapre sentation stream. Slides from the same presentation typically have similarcolorhistograms,dominantcolors,etc. Th erefore, most color features are not strong discriminators i n these

images.Ontheotherhand,slidesinapresentation contain different text, combinations of text lines, layouts , images, and graphics. Our experiments showed that the text contentandthelayoutofthetextlinesofaslide arestrong discriminatory features for slide images. Therefore , we utilizethetextlayoutandthetextcontentofsli deimages tolinkthepresentationstreamstodigitalcamera imagesof slides.Ageneralflowofouralgorithmisshownin Figure 2.Thenextsectionsexplainhowweextractandmat chtext linelayoutandtextinslideimages.

3.1 TextLayoutExtraction/Matching

Text regions in images contain strong horizontal an d verticaledges.FirstaSobeloperatorappliedtot heimage to obtain edge magnitudes. After edge magnitudes ar e computed for an image, an edge is detected only if the edge magnitude is larger than a threshold. Because the background/foreground contrast in a slide image may be obtained by color contrast as well as luminance con trast, edges are detected using R, G, B, and Luminance col or components separately. An edge pixel is detected if an edgepixelisidentifiedinanyofthecolorspaces .Foreach pixel location, a value, E_{xy}, is computed by accumulating thenumberofedgepixelsinaneighborhoodwindow K×L. The pixels that have E_{xy} values that are larger than an adaptive threshold are marked as pixels that belong to a textregion.Connectedcomponentanalysisandaspec tratio informationisthenusedtoidentifythetextregio ns.Figure 3 shows the identified text regions from a digital camera

slide image. The figure also shows that the edge de tection results are significantly better when all the color components are used for edge detection versus using just the luminance component.
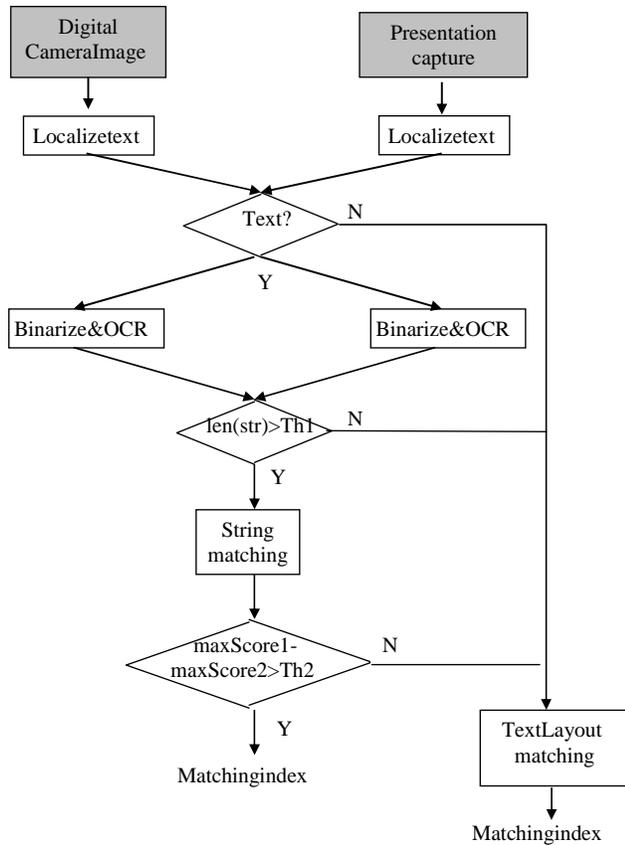

Figure 2. Slide matching with text and layout

After text regions are identified, text line profil e is extracted by computing, for each horizontal line in the image, the maximum run of pixels that belong to a t ext region. In order to compute the similarity distance between the text line profiles of two slide images, first t he text line profiles are normalized in both horizontal and vert ical directions using the highest peak. Note that this h igh peak corresponds to the longest text line in the slide i mage. The distance between two line profiles are found by ali gning the profiles with their peak value and computing th e sum of absolute differences.

3.2 String Extraction/Matching

After text regions are identified, a commercial OCR package [5][6] is used to extract text from slide i mages. OCR packages usually correct for small skew and rotations, therefore we do not address these proble ms here.

After the images are OCR'd, string matching is perf ormed on the text output to find a similarity score betwe en two slide images. OCR results from slide images capture d by different devices can vary widely. For example, th e text output extracted from a digital camera image is gen erally less accurate than that obtained by OCR'ing the scr een projection output for the same slide. In most appli cations one of the capture sources is likely to be more rel iable than the other and the OCR results obtained from one of the sources could be close to the ground truth. We take this into consideration when performing string matching and define the string that is obtained from the more re liable source as the ground truth string. The characters obtained for each slide are first concatenated. Then the si milarity score between two strings are computed as $d_s = (\ell_g - d_e)/\ell_g$, where ℓ_g is the length of ground truth string and d_e is the edit distance between two strings.

Edit distance is computed by counting the number of insertions and deletions required for matching. The punctuations and the extra characters in the string that is extracted from the less reliable source are ignored .

4. EXPERIMENTAL RESULTS

In this section, we present experimental results on the retrieval performance obtained by our proposed algo rithm. In the first experimental setup, image matching is performed on the images obtained from the following two sources:

- Presentation slides in PowerPoint that are saved as a series of JPEG images.
- Digital camera: 2-4 MPixel images mostly containing the slide region, with rotation less than ±5 degrees.

Here, we performed experiments on a presentation recording database containing 41 presentations with 1160 slides. We collected 109 digital camera images to b e used for queries in our experiments during 7 different presentations given in our lab. Text extraction is performed on the PowerPoint slides using the Microsoft APIs a t the time they are inserted into the database. All the d igital camera images contain at least one line of text. Th ese images vary in terms of room lighting, use of flash , distance to the projector, motion blur, occlusion, etc. The process of matching one digital image to a database of 1160 images was performed in under 3 seconds on a 1 GHz PC. Note that segmentation of the slide region from the digital camera image is not required for the matchi ng techniques employed here.
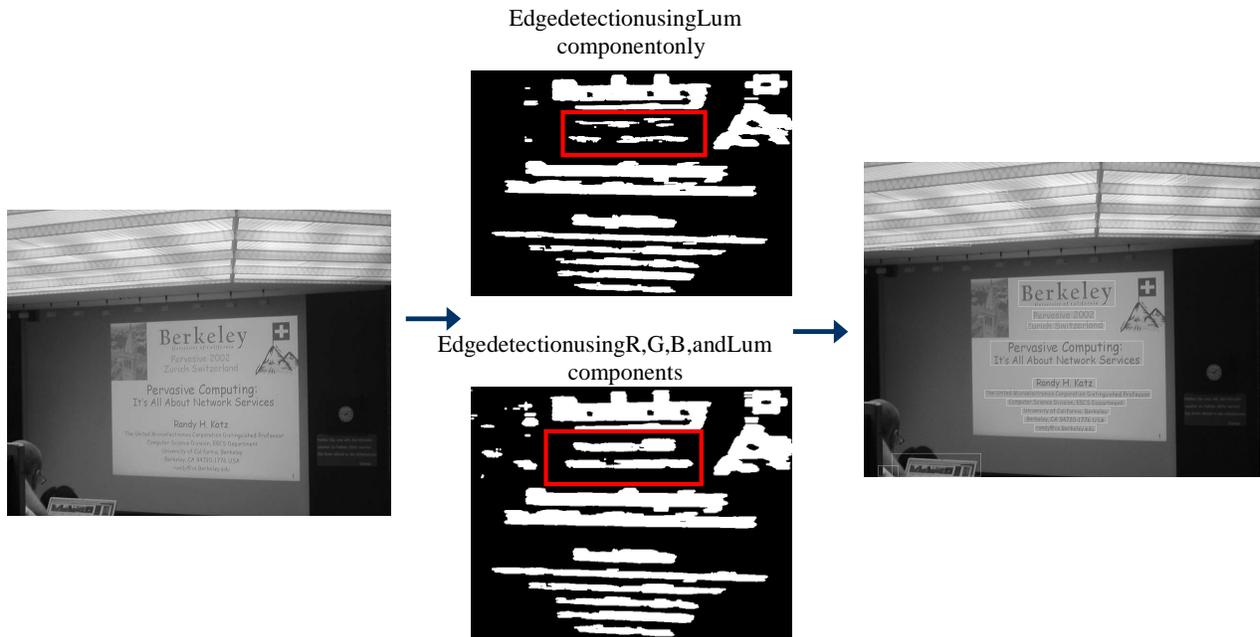
Figure3.Textlocalizationdetectioninadigital   cameraimage.

We also ran another set of experiments where the sl   ide images captured by the *Presentation Recorder* are retrieved by using the digital camera images. In th   is case, the database contained 41 presentations with 4814 captured presentation recorder images (i.e. screen   capture). Most of these images were presentation slides. Some   of these images have the same slide content because a presenter may visit a slide more than once. For eac   h query, the 10 slide images with the highest similarity sco   re are retrieved.

The retrieval results obtained by querying the pres   entation slides with digital camera images are presented in   Table 1. The retrieval rate is defined as the ratio of corre   ctly retrieved relevant images to the number of relevant   images in the database. The ground truth images are the presentation slides that match the digital camera i   mages in terms of their slide content. As can be seen from T   able 1, the average retrieval rate (recall) obtained by que   rying the PowerPoint slide database is 95%. The 3 [rd] presentation, where the presentation slides generally contained v   ery few lines of text, yielded the lowest retrieval rate. T   he average retrieval rate obtained using the Presentation Reco   rder database is 87%. Some of the presentations containe   d slides with small fonts (e.g., 2 [nd] and 4 [th] presentations). In these cases, the OCR on the slides captured by the PresentationRecorder yielded low accuracy, resulti   ng in a lower retrieval performance. However, the retrieval performance using the PowerPoint slides was not aff   ected

by this since in this case we extract the text dire   ctly from the PowerPoint file instead of using OCR.

Table1.Retrievalresultswhendigitalcameraimag   es areusedtoquerypresentations.

Pres. no	Number of digital camera images	AveRetrievalRate (PowerPoint slides)	AveRetrieval Rate(Presentation Recorderslides)
1	19	0.94	0.91
2	7	1.00	0.75
3	11	0.91	0.91
4	9	1.00	0.85
5	20	0.95	0.86
6	22	0.96	0.92
7	21	0.95	0.84
ALL	**109**	**0.95**	**0.87**

One should keep in mind that the overall retrieval   accuracy of this method greatly depends on the quality (reso   lution, sharpness, etc.) of the digital camera image as muc   h as the amount of text content in the captured presentation   slide. Note that it is possible to improve the retrieval a   ccuracy of slide image retrieval by incorporating information   from the digital camera image timestamps.
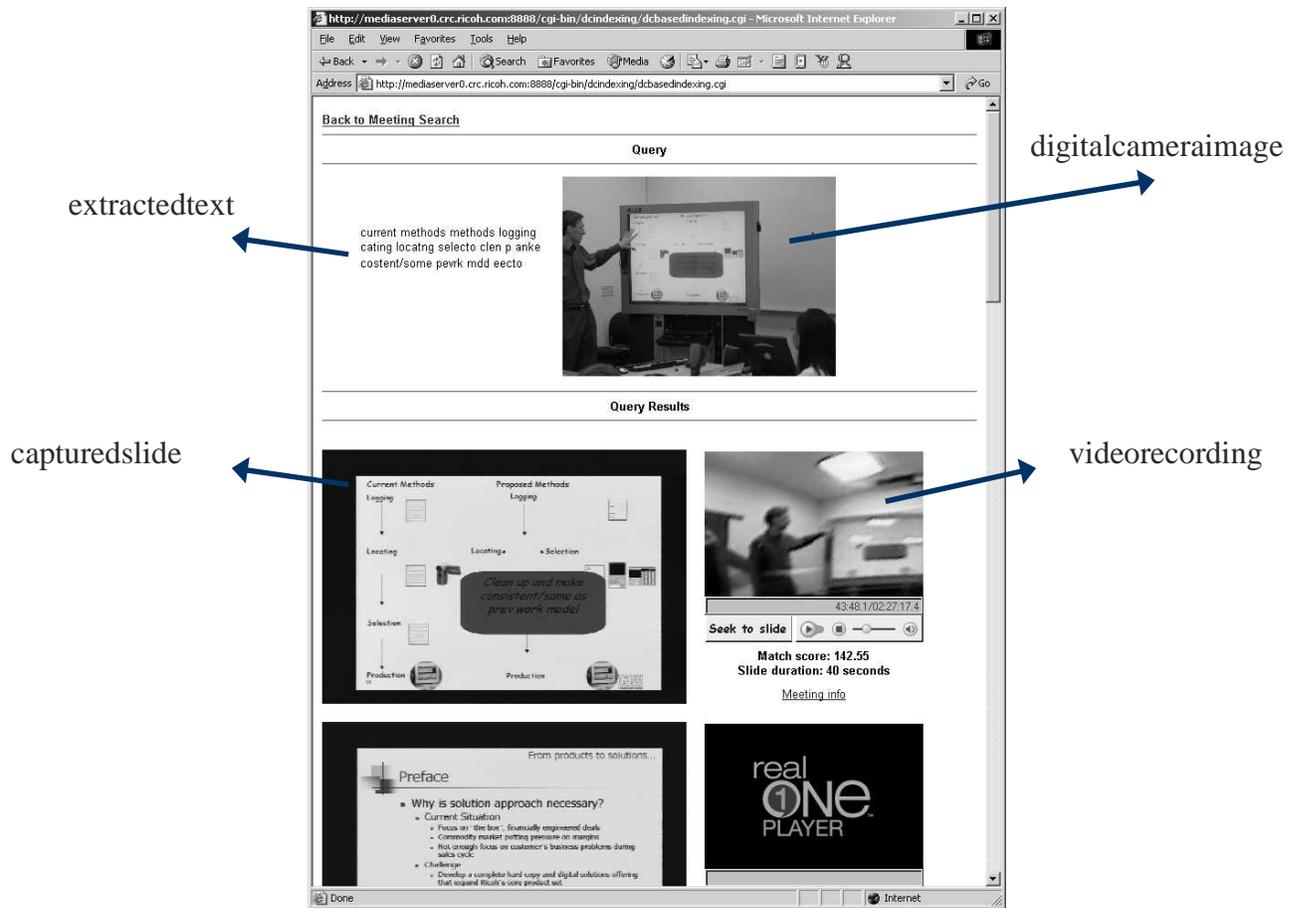
extractedtext

digitalcameraimage

capturedslide

videorecording

Figure4.Digitalcamera-basedretrievalinterface   forpresentationrecordings.

Theretrievalresultsarepresentedintheinterfac   eshownin Figure 4. For a given query image, the matched presentation slides and their corresponding audiovi   sual recordingsaredisplayedinanorderbasedonthes   imilarity score. User then playback the video recording of th   e presentertalkingabouttheparticularslidethath   ecaptured withthedigitalcameraimage.

5. CONCLUSIONS

In this paper, we presented a novel retrieval appli   cation and a novel method for linking multimedia streams captured during a presentation. Linking presentatio   n streams and documents through slide image matching opensuppossibilitiesfornewwaystoindex,retri   eve,and access meeting/ presentation/ lecture content and, consequently,formanyinterestingandusefulappli   cations.

6. REFERENCES

[1] Mukhopadhyay, S., and Smith, B., "Passive capture and structuring of lectures", ACM Multimedia pp. 477-487,1999.

[2] Chiu,P.,Foote,J.,Girgensohn,A..andBoreczky,   J., "Automatically Linking Multimedia Meeting Documents by Image Matching" proceedings of Hypertext'00,ACMPress,pp.244-245,2000.

[3] Chiu,P.,Kapuskar,A.,Wilcox,L.,"MeetingCaptur   e in a Media Enriched Conference Room," 2nd International Workshop on Cooperative Buildings, pp.79-88,1999.

[4] Lee,D.S,Erol,B.,Graham,J.,Hull,J.J.,andMur   ata, N., "Portable Meeting Recorder", ACM Multimedia Conference,pp.493-502,2002.

[5] Scansoft Capture Development System, www.scansoft.com

[6] TransymOCRengine,http://www.transym.com/