Integrating Computer Vision with Web Based Knowledge for Medical Diagnostic Assistance

Aviv Segev

KAIST - Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea

aviv@kaist.edu

Abstract

The analysis of medical documents necessitates context recognition for diverse purposes such as classification, performance analysis, and decision making. Traditional methods of context recognition have focused on the textual part of documents. Images, however, provide a rich source of information that can support the context recognition process. The paper proposes a method for integrating computer vision in context recognition using the Web as a knowledge base. The method is implemented on medical case studies to determine main symptoms or achieve possible diagnoses. In experiments the method for integrating computer vision in context recognition achieves better results than TF/IDF and only context recognition. The proposed method can serve as a basis for an image and text based decision support system to assist the physician in reviewing medical records.

Keywords: Computer-assisted diagnosis; Computer vision; Decision support system

1 Introduction

The field of document analysis requires context recognition for classification, performance analysis, and decision making. Traditional methods of context recognition focus on the textual part of documents. However, images constitute a rich source of information that can complement the context recognition process and computer vision techniques can supply information about images for diverse uses.

In fields which rely on extensive information for decision making, such as medicine, information is extracted from multiple sources. Most of the information consists of both textual and image formats. An analysis that relies on a single format of information, such as text, may be deficient. Therefore, methods that support the automatization of the process of computer-assisted medical analysis require the integration of both textual analysis techniques and computer vision techniques.

As an example of the variety of information a physician encounters when analyzing a patient, part of a case study taken from Clinico-Pathological Conference is displayed in Figure 1. The medical case studies in the Clinico-Pathological Conference of the Johns Hopkins University School of Medicine consist of text and images and present the clinical course, radiological findings, and relevant laboratory results for a particular patient and present the medical analysis [1].

To analyze information consisting of both text and images a model of the integration of both methods is described in Figure 2. The input is separated into text and image. The next step implements a context recognition model for textual analysis and a computer vision model for image analysis. Then the vision is integrated into context, yielding conceptual output. For example, in the field of medicine, the model input can be a medical case study and the model output is a list of words that represent major symptoms or possible diagnoses and these words are checked against the solutions in the medical case studies.

The next section describes the related work, and Section 3 presents the model and processing. Section 4 analyzes experiments of the model in the field of medicine. Section 5 discusses the differences between the methods in the experiments and possible implementations. Finally, the conclusion and further research are described in Section 6.

JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE

CLINICO-PATHOLOGICAL CONFERENCE

Case Study
CPC #4: Tuesday, December 17, 2002

Chief Complaint:
A 46-year-old male with nausea and vomiting status post heart transplant who collapsed in his bathroom.

History of Present Illness:
The patient was a 46-year-old male with a long history of diabetes mellitus type II and multiple myocardial infarctions between 1993 and 2000. In July 2001 he was diagnosed with ischemic dilated cardiomyopathy. In August 2001 he was admitted for an acute exacerbation of congestive heart failure and received a left ventricular assist device. A cardiac biopsy at the time demonstrated severe hypertrophy and interstitial fibrosis consistent with a transmural healed infarct (Figure 1). Shortly thereafter, he developed acute renal failure requiring continuous veno-venous hemodialysis and also suffered a cerebrovascular accident in the watershed areas of the frontal and parietal cortex (left>right).

Figure 1

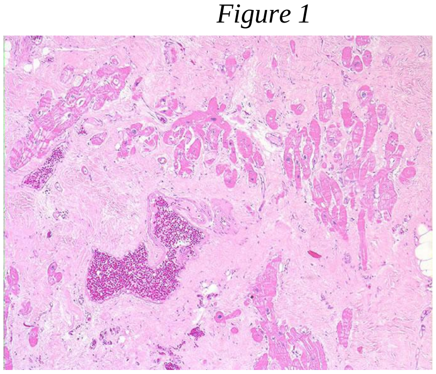
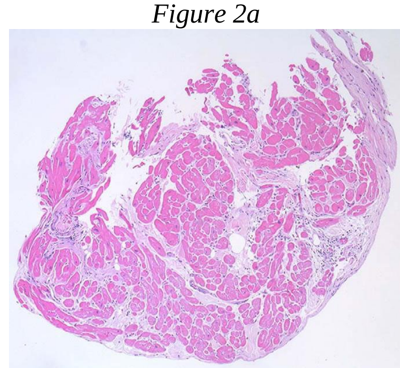


Figure 2a



• • •

What is the most likely cause of the patient's demise?

Figure 1: Sample Case Study

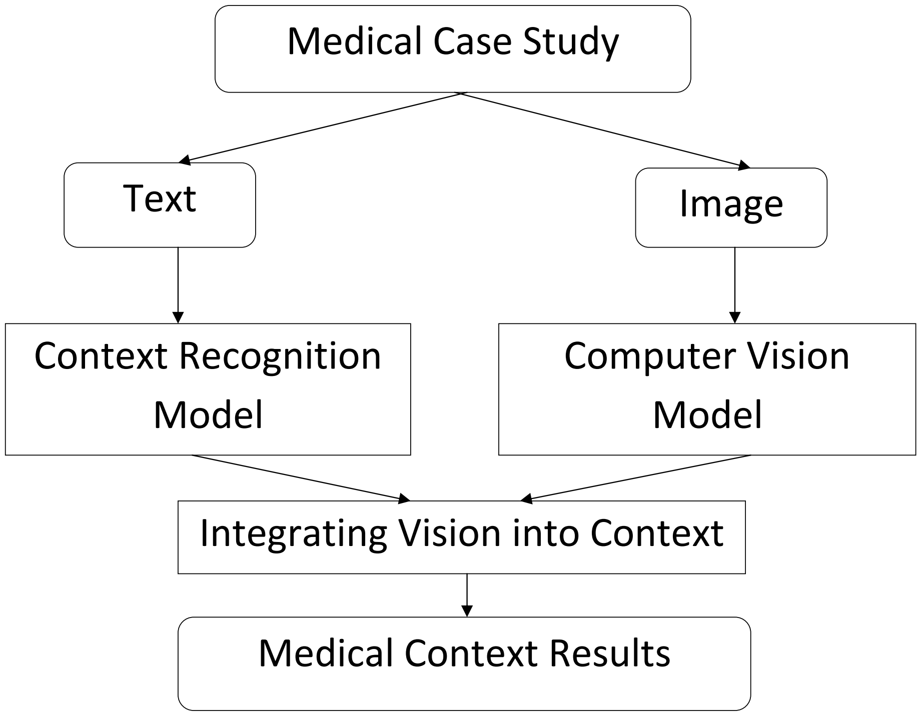


Figure 2: Method Outline

2 Related Work

Previous work explored the interaction of textual and photographic information in document understanding. The use of complementary information in scene understanding has been explored in computer vision systems that use scene context in the task of object identification. Work has been done on extracting picture-specific information from text accompanying a photograph [14]. In addition, a method was proposed for creating web annotations in a controlled natural language to express domain-specific ontological knowledge about that website in an unambiguous subset of English [11]. Previous work also included an ontology definition language for agents working on the semantic web [6].

Another application is a model of object recognition as machine translation described in [5]. In this model, recognition is a process of annotating image regions with words. First, images are segmented into regions, which are classified into region types using a variety of features. Then, a mapping between region types and keywords supplied with the images is learned.

Word sense ambiguity problems are addressed in [2]. The approach bases on a method for automatically annotating images by using a statistical model for the joint probability for image regions and words. The model is learned from a database of images with associated text. To use the model for word sense disambiguation, the predicted words are constrained to be possible senses for the word considered.

The problem of modeling annotated data, data with multiple types such as image and text captions, is addressed in [4]. The work describes three hierarchical probabilistic mixture models that aim to describe such data, culminating in correspondence latent Dirichlet allocation, a latent variable model effective at modeling the joint distribution of both types and the conditional distribution of the annotation given the primary type.

Similarly, computer vision can provide information for integration with other systems. The integration of speech and image using Bayesian networks is found in [17]. Object recognition errors are taken into account by conditional probabilities estimated on test sets. The Bayesian network is dynamically built up from verbal object description and is evaluated by an inference technique combining bucket elimination and conditioning.

Recent work in the field of medical knowledge acquisition includes a language-independent approach for extracting knowledge from medical natural language documents [16]. Another work compared different classification algorithms in clinical decision-making [15]. In addition, probability for case-based reasoning applied to the field of medical diagnosis was presented in [9].

A method of medical diagnostic assistance supported solely by text-based context analysis was described in [13]. In this paper, a method for integrating text based context recognition and computer vision using the Web is presented. The proposed method improves context recognition, based on both non-structured textual analysis and semi-structured computer vision data. The proposed model integrating computer vision and context recognition takes a document from the Web containing text and pictures and returns a set of words representing the document context.

3 Model and Processing

This section presents the proposed Web-based model for the integration of context recognition and computer vision. First, the two main components are presented, namely the context recognition model for the detection of possible topics and the computer vision model for the definition of images. Then the integration of the two is described: the image related information is compared with the generated contexts and then with the text of the document for further verification.

3.1 The Context Recognition Model

Several methods have been proposed in the literature for extracting context from text. A set of algorithms was proposed in the Information Retrieval community, based on the principle of counting the number of appearances of each word in the text, assuming that words with the highest number of appearances serve as the context. Variations on this simple mechanism involve methods for identifying the relevance of words to a domain, using methods such as stop-lists and term frequency and inverse document frequency (TF/IDF) [10].

The paper integrates a model of context recognition using the Web as a knowledge base, thus giving a context recognition model immediate access to a nearly infinite amount of data in a multiplicity of fields [12]. Context is represented as a set of descriptors and a set of weights to describe a given situation. The model does not require large training sets and since the Web is used as the database, there is no database maintenance.

The context recognition model used is based on the definition of context as first class objects formulated by McCarthy [8]. McCarthy defines a relation $ist(C, P)$, asserting that a proposition P is true in a context

C. This relation is used when discussing context extraction.

A *context* $\mathcal{C} = \left\{ \{ \langle c_{ij}, w_{ij} \rangle \}_j \right\}_i$ is a set of finite set of descriptors c_{ij} from a domain D with appropriate weights w_{ij} that define the importance of c_{ij}. For example, a context C may be a set of words (and hence, D is a set of all possible character combinations) defining a document Doc, and the weights could then represent the relevance of a descriptor to Doc.

Let $P_1, ..., P_m$ be a series of textual propositions representing a document, when $\forall P_i$ there exists a collection of sets of contexts C_{ij} so that: For each i, $ist(C_{ij}, P_i) \forall j$ meaning that the textual proposition P_i is true in each of the set of contexts C_{ij}. The context recognition algorithm [12] identifies the outer context set C defined by

$$ist(\mathcal{C}, \bigcap_{i=1}^{m} ist(\mathcal{C}_i, P_i)) \forall j.$$

The input to the algorithm is a stream, in text format, of information. The context recognition algorithm output is a set of contexts that attempts to describe the current scenario most accurately. The set of contexts is a list of words or phrases, each describing an aspect of the scenario. The algorithm attempts to reach results similar to those achieved by the human process of determining the set of contexts that describe the current scenario.

The context recognition algorithm consists of the following major phases: collecting data, selecting contexts for each text, ranking the contexts, identifying the current contexts, and obtaining the multiple contexts.

- Collecting Data - The information from the information sources is decomposed into words and the keywords are extracted from them.

- Selecting Contexts for Each Text (Descriptors) - For each keyword a set of preliminary contexts is extracted from the Web, which is used as a context database.

- Ranking the Contexts - Each preliminary context is ranked according to the number of references it receives in the context database and the number of appearances it has in the text.

- Identifying the Current Contexts - The preliminary contexts that have significantly higher numbers of references and higher numbers of appearances are included in the current set of contexts.

- Obtaining the Multiple Contexts - The current contexts are examined for synonyms and synonymous contexts are united.

The success of the algorithm depends, to a great extent, on the number of documents retrieved from the Web. With more relevant documents, less preprocessing (using methods such as Natural Language Processing) is needed in the data collection phase.

3.2 The Computer Vision Model

The field of computer vision explores automatic pattern or object recognition. Methods are developed to discover which features distinguish objects and to design classification algorithms. The concepts of feature extraction that respond to characteristic changes in brightness, color, and texture associated with natural boundaries, used in the computer vision method integrated in the present model, are based on the concepts of Martin et al. [7].

Computer vision models usually compute a set of real-valued features that represent visual properties such as size, position, color, and texture. For example, consider in the field of medicine a simple computer vision model as proof of concept. The computer vision model analyzes four types of images: blood, brain, chest, and liver.

- *Blood* - Blood is recognized by parsing the image into three color based segments and checking whether one is similar to blood color. This yields close to perfect recognition of blood images.
- *Brain* - The brain is identified in two steps. First, the brain is recognized by checking roundness of the biggest object in the image. Then, the shadowed image of the main object in the picture is covered by a brain shadow and the difference is examined. The brain recognition method yields relatively low accuracy results of 37.1%.
- *Chest* - The method of the identification of the chest and liver is based on locating the bones, which are the lightest objects in the picture. The biggest white object is checked to see whether it matches the location (bottom center) and is proportional in size to the image of a vertebra.
- *Liver* - Once the vertebra is located, the method searches for the front rib bones. If the front rib bones are located, then the image is identified as chest. Otherwise, it is identified as liver.

The result of the Algorithm 1 is the identification of one of the above body parts.

Algorithm 1 Simple Vision Model (Image)

```
if color image then
  if contains blood color then
    identify blood
  end if
else
  if locate brain then
    identify brain
  else
    if locate vertebra then
      if front ribs then
        identify chest
      else
        identify liver
      end if
    end if
  end if
end if
```

3.3 Integrating the Vision into the Context

The context recognition model and the computer vision model are integrated according to the following algorithm (Algorithm 2).

- Each possible image X_i has prior knowledge about identification failure rate (X_{i1}, X_{i2}) according to previous domain knowledge. X_{i1} defines the false negative: the algorithm fails to identify the image. X_{i2} defines the false positive: the algorithm erroneously identifies an image.
- For each domain for all contexts prior knowledge about identification failure rate (Y_1, Y_2) exists and is similarly defined.
- For each context recognition result (CRR) compare to computer vision result (CVR). The algorithm searches for the possible results of the integration in the text. Once the possible contexts are defined, they can be checked and verified against a local domain knowledge base. This was performed in the examination of medical case studies.

Algorithm 2 Integration Algorithm (CRR, CVR)

if CVR = CRR **then**
  Verify against domain information
  Add context with probability
  $0.5(1 - X_{i1})(1 - Y_1) + 0.5(1 - X_{i2})(1 - Y_2)$
else
  Search text (T) for CVR or dictionary synonym (DS)
  if T = CVR or DS = CVR **then**
    Verify against domain knowledge and
    rank CVR as context with probability
    $0.5(1 - X_{i1})Y_1 + 0.5X_{i2}(1 - Y_2)$
  else
    Add CVR as context with probability
    $0.5X_{i1}(1 - Y_1) + 0.5(1 - X_{i2})Y_2$
  end if
end if

3.4 Medical Case Example

An example of a sample test case is presented in Figure 1. The results of the computer vision identified the images in the case study as `Blood`. The results of the text included the following possible diagnosis of: `blockage or defect of artery`, `heart attack`, `coronary artery disease (CAD)`, `valvular or congenital disease`, and `other cardiac disorders`.

The results are checked against the medical information lists constructed from the Merck Manual in a process of 'reverse' classification in which the types of images and body parts are matched with the types of diagnoses. Then, according to these medical information lists, presented in Table 1, the results represent main symptoms or possible diagnoses. Then the algorithm again searches in the text for possible contexts from the list of diagnoses in the table. The word `heart` appears since the patient had heart failure and a previous heart attack. Similarly, the patient has suffered from `coronary artery disease (CAD)`. The word cardiac appears three times in the text.

The answer to the medical case study states that the patient expired from Coronary Allograft Vasculopathy (Accelerated Graft Arteriosclerosis), which resulted in cardiogenic shock. The algorithm yielded heart and artery related results, displaying the advantage of the integration of the images and the web based contexts in assisting in the analysis of case studies.

Table 1: Possible Diagnosis Based on Integration of Image and Text

Image Type	Body Area Tested	Procedure	Possible Diagnosis
X-ray	Any artery in the body; commonly in brain, heart, kidneys, aorta or legs	Arteriography (angiography)	blockage or defect of artery
X-ray	Liver	Percutaneous transhepatic cholangiography	Tumor
X-ray	Veins	Venography	Blockage of vein
Ultrasound	Liver	Ultrasonography (ultrasound scanning)	Tumor, Jaundice
Positron emission tomography (PET)	Brain	Radioactive imaging to detect abnormality of function	Epilepsy, brain tumor, stroke
Radionactive Radionuclide imaging	Many organs	Radioactive imaging to detect abnormality of blood flow, structure, or function	Heart attack, coronary artery disease (CAD), valvular / congenitial disease, and other cardiac disorders

4 Experiments

Medical case studies are used to test the model performance. Domain information is obtained from the Merck Manual [3], which provides medical information lists for the classification of possible diseases and diagnostic procedures using imagery techniques such as X-ray, CT, Radionactive imaging, and microscopic examination. This information was organized to correspond to the present method of image and text integration. Table 1 displays a sample of possible diagnoses based on the image and textual information.

4.1 Data

The data for the experiments were taken from existing medical cases from different sources. The experiments included 92 medical cases collected from sources including Web sites from Johns Hopkins University, University of Wisconsin-Oshkosh, Southern Illinois University, University of Texas Medical Branch, Rutgers University, University of New England, Weber State University, Vanderbilt Medical Center, Nature Journal, Medscape, Cedars-Sinai Medical Center, University of Virginia, Washington University, and Harvard University. All the medical cases included both textual descriptions and images.

The total number of images in all the cases was 365. The images can be categorized according to the following classification: `Abdomen`, `Biopsy`, `Blood`, `Brain`, `Chest`, `Liver`, `Lung`, `Renal`, and `Other`.

4.2 Experiments Process

The textual part of the medical records is fed into the context recognition model and processed. A list of possible contexts, which can include possible diagnoses and main symptoms, is obtained. The images from the case are analyzed by the computer vision model. Next, according to the algorithm, the results of the context recognition and of the computer vision are compared. If the results match, they are checked against the medical information lists constructed from the Merck Manual. Then, according to the medical information presented in Table 1, the results represent main symptoms or possible diagnoses. Each diagnosis receives a probability according to Algorithm 2. If the results of the computer vision and the context recognition do not match, then a search is performed on the case study text for words that match the results of the computer vision. If no match is found, then the results of the two processes are weighted according to Algorithm 2.

The Integration Algorithm (Algorithm 2) requires prior knowledge of the failure rate probability of both

Table 2: Computer Vision Algorithm Data

Image Type	Sample Set Size	Training Set Size
Blood	61	8
Brain	68	6
Chest	27	-
Liver	37	-
Other	31	-

the context recognition model and the computer vision model. The prior knowledge regarding the failure rate of the context recognition model is based on prior work of implementing the context recognition model on medical cases [13]. According to the prior results, the false negative rate (Y1) are 30% and the false positive rate (Y2) are 54.5%. The prior knowledge failure rate of the computer vision model is based on sample testing for each of the identification possibilities described in Simple Vision Model (Algorithm 1). Since some case studies include a large number of the images, a subset of 224 images was selected from the images to represent the cases more accurately. A training set for identifying `Blood`, `Brain`, `Chest` and `Liver` was used to calibrate the image processing algorithm. The prior identification rate for the image processing was set to perfect recognition 100% to represent the result of the training cases.

The image processing experiments results, including the achieved true positives and false positives, are described next, along with a comparison of the performance of the integrated image and context model to the performance of previously existing models.

4.3 Results

The section first analyzes the results of the computer vision algorithm separately, followed by an analysis of the integrated model performance using both Web context recognition and computer vision algorithms. Table 2 presents the computer vision data analyzed according to the following breakdown. The sample set included 61 images of `Blood`, 68 images of `Brain`, 27 images of `Chest`, 37 images of `Liver`, and 31 images not defined in any of the categories and classified as `Other`. From each sample set a training set was selected to calibrate the computer vision model. The number of images used for training was 8 for `Blood` and 6 for `Brain`. For the other images no training data were required for the algorithm.

The results can be analyzed as a two-class prediction problem (binary classification), in which the outcomes are labeled either as positive (p) or negative (n) class. A binary classifier yields four possible outcomes. If the outcome from a prediction is p and the actual value is also p, then it is called a true positive (TP); but if the actual value is n then it is called a false positive (FP). Conversely, a true negative occurs when both the prediction outcome and the actual value are n, and false negative (FN) is when the prediction outcome is n while the actual value is p.

To draw a Receiver Operating Characteristic (ROC) curve, only the True Positive Rate (TPR) and False Positive Rate (FPR) are needed. TPR determines a classifier or a diagnostic test performance on classifying positive instances correctly among all positive samples available during the test. FPR, however, defines how many incorrect positive results occur among all negative samples available during the test. The TPR and FPR are defined by:

$TPR = TP/(TP + FN)$

$FPR = FP/(FP + TN)$

The best possible prediction method would yield a point in the upper left corner of the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). The top left point is also called a perfect classification. A completely random guess would give a point along a diagonal line from the left bottom to the top right corners. The diagonal line divides the ROC space in areas of good or bad classification/diagnostic [18].

The results of the computer vision algorithm are presented in Figure 3, which displays the ROC curve. The X-axis represents the False Positive Rate and the Y-axis represents the True Positive Rate. The results of the computer vision `Blood` identification were high, displaying 96.23% true positive identification and only 3.77% false positive. These results emphasize the advantage of using the simple algorithm based on identifying the color spectrum of the red parts of the image for identifying a blood image. The `Chest` identification results were also high, achieving 88.89% true positive and 17.24% false positive. The Liver results were less successful, with 78.38% true positive and 46.30% false positive. The least successful model was the identification of the `Brain` with only 37.10% true positive and 20.69% false positive. The high value of false positive of the `Other` category can be attributed to the misclassification of the images from the rest of the image types.

To evaluate the performance of the integrated computer vision with Web based context algorithm, the performance of two common algorithms was analyzed based on the same data. The first method is based on term frequency and inverse document frequency (TF/IDF) [10]. Each case was processed using the TF/IDF algorithm and a threshold which filters out words with a frequency count higher than the second standard

Figure 3: Computer Vision Algorithm Identification Results By Image Type

deviation from the average frequency of a word within a document. To analyze the ROC curve an additional result of all the TF/IDF words was also evaluated, which is equivalent to analyzing all the words in the document.

The second method analyzed was the Web based context recognition algorithm described in Section 3.1, without the computer vision component. For each case study evaluated, three degrees of results were analyzed: the top ranking contexts, the top 20% of all ranked contexts, and the full list of all possible contexts proposed by the algorithm.

The third method analyzed was the integrated computer vision with Web based context algorithm described in Section 3.3. The method was analyzed similarly to the Web based context algorithm with the three degrees of results, where each level integrates both vision results and the Web based results.

All the methods were first analyzed for diagnoses that include the identification of the correct diagnosis. In other words, a true positive would be defined if the diagnosis appears in the list of outputs of the algorithm. A false positive is defined if in the algorithm output there is a list of diagnoses which do not include the real diagnosis. A false negative is defined as no diagnosis at all.

The results of the Receiver Operating Characteristic (ROC) curve for all three methods diagnoses that

Table 3: Area Under Curve (AUC) - Method Comparison of Diagnoses that Include the Identification of the Correct Diagnosis

Method	TF/IDF	Context	Image + Context
AUC	0.629	0.754	0.808

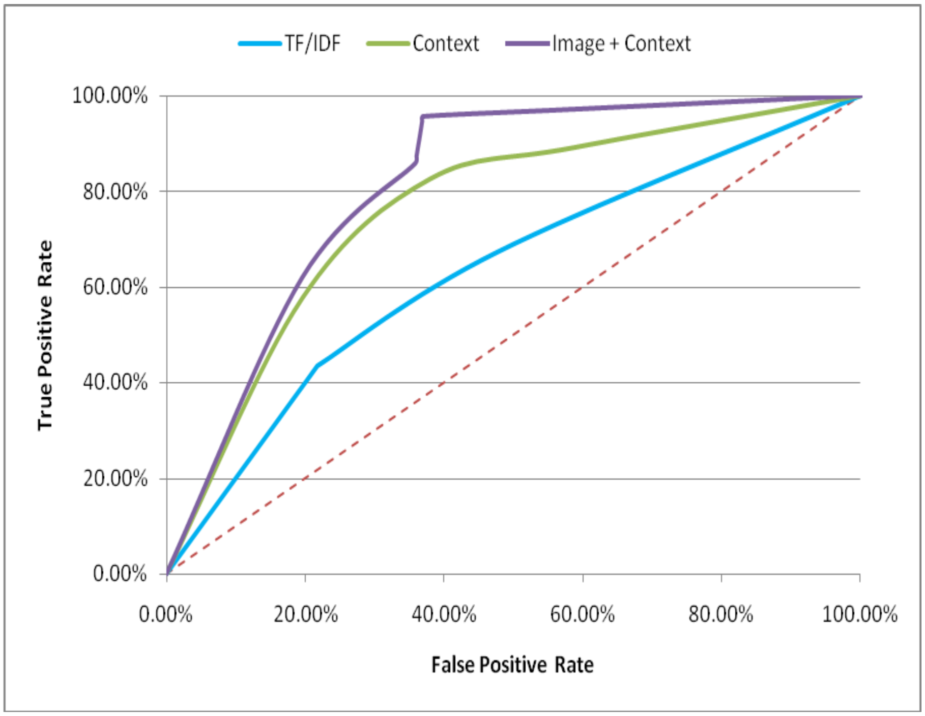


Figure 4: Comparison of Methods for Identification of Diagnoses that Include the Correct Diagnosis

include the identification of the correct diagnosis are displayed in Figure 4. The X-axis represents the false positive and the Y-axis represents the true positive. The three graphs represent the three methods: TF/IDF, Web context (Context), and integrated computer vision and Web context (Image + Context). The integrated computer vision and Web context method dominates the other methods. The performance of the Web context method also dominates the TF/IDF method by a large margin.

The evaluation of the Area Under Curve (AUC) for the diagnoses that include the identification of the correct diagnosis displayed in Figure 4 appears in Table 3. According to the results, the TF/IDF AUC is 0.629, which is considered a poor performance, the Web context based method achieved AUC of 0.754, which is evaluated as a fair identification result, and the integrated computer vision and Web context resulted in AUC of 0.808, which is considered a good identification result.

The methods were next analyzed for symptom identification for correct diagnosis. In this case a true

Table 4: Area Under Curve (AUC) - Method Comparison of Symptom Identification for Correct Diagnosis

Method	TF/IDF	Context	Image + Context
AUC	0.673	0.844	0.895

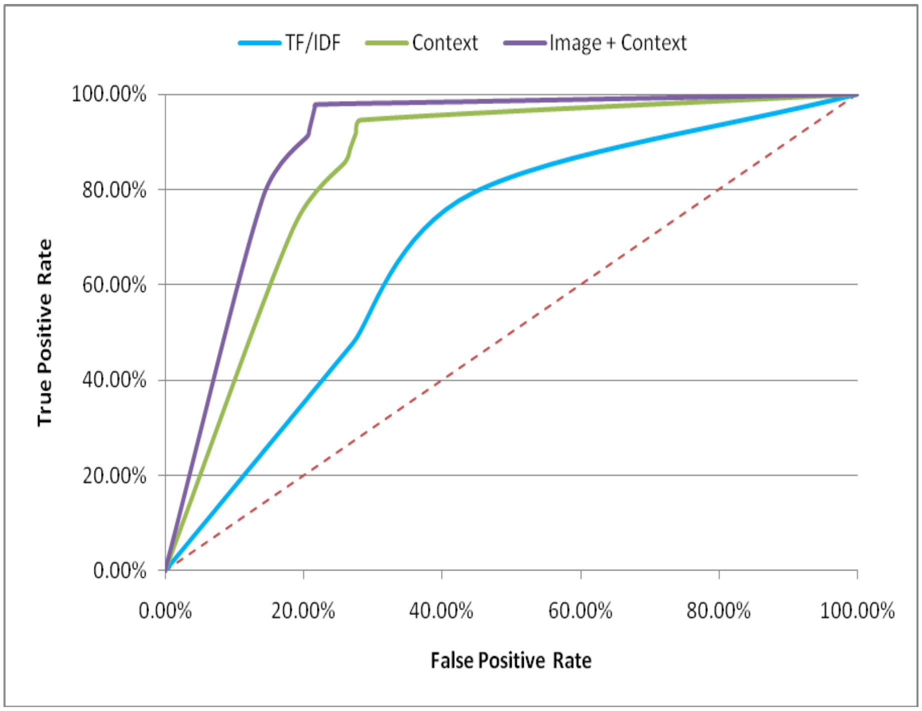


Figure 5: Comparison of Methods for Identification of Symptoms Relevant to the Correct Diagnosis

positive would be defined as any symptom which appears in the algorithm list of outputs and is related to the correct diagnosis. A false positive is defined as a list of symptoms irrelevant to the correct diagnosis and a false negative is defined as no symptoms at all.

The results of the Receiver Operating Characteristic (ROC) curve for all three methods symptom identification for correct diagnosis are displayed in Figure 5. The computer vision and Web context method again dominates all other methods, followed by the Web context method, with the TF/IDF achieving the lowest results. The results here are higher since the correct diagnosis results include in most cases also the relevant symptom identification.

The evaluation of the Area Under Curve (AUC) for the diagnoses that include the identification of the correct diagnosis displayed in Figure 5 appears in Table 4. According to the results the TF/IDF AUC is 0.673, which is considered a poor performance, the Web context based method achieved AUC of 0.844, which is evaluated as a good identification result, and the integrated computer vision and Web context resulted in AUC of 0.895, which is considered a good identification result, bordering on an excellent identification result (AUC $\geq$ 0.9).

The reasons for the difference in performance between the algorithms and possible implementations of the integrated computer vision and Web context are discussed next.

5 Discussion

The main advantage of both the Web context method and the integrated computer vision and Web context over the TF/IDF is the ability to identify a symptom or cause of death which does not appear in the text itself. While the latter has to work within the limits of the original case study text, the context analysis method goes out to the Web, using it as an external judge and returning keywords that are deemed relevant, although they were not originally specified in the case description.

The advantage of the integrated computer vision and Web context model compared to the Web context model can be seen in the reduction of the false positive results. Although the Web context by itself in most cases returns the correct results, the ranking of the result is not always high in the result list. The computer vision results allow the identification of which context results should receive higher ranking and consequently the model identifies the correct diagnosis or relevant symptom.

One of the problems identified with the TF/IDF model is that it overlooks some topics which include simple explanations such as in case studies describing `Blood` or `Brain`. The term is not included in the TF/IDF results due to IDF - the term appears too much in other documents and as a result does not appear as relevant in the algorithm results.

The model achieved high results in both identifying diagnoses that include the identification of the correct diagnosis and identifying symptoms for correct diagnosis. A possible implementation of the model could include a decision support system for a physician analyzing a case. Alternatively, an implementation of the model could be used as a second opinion tool for the patient or his family. Since the model in most cases supplies a list of diagnoses, including the correct diagnosis, a physician would be able to receive an extended list and rule out the incorrect diagnoses.

6 Conclusions and Further Research

The paper presents a Web-based technique of integrating context recognition and computer vision and demonstrates how this method can be implemented. The paper uses real medical case studies to experiment with the proposed method.

Usually document analysis focuses on the text part of a document, but this work proposes an idea of text understanding by understanding image first, since image can constitute a rich source of information. This idea is based on the assumption that the accuracy of computer vision is high enough to provide a useful hint for context recognition, since an inaccurate computer vision system might also mislead the overall context recognition.

The findings show that the proposed integration method yields improved results in comparison to the separate use of context recognition or TF/IDF methods. Additionally, use of state-of-the-art as opposed to simple computer vision algorithms can improve the results.

The main advantage of the proposed model for the integration of computer vision into context recognition is its use of the Web as a knowledge base for data extraction. The information provided by the computer vision model complements and augments the context recognition process by reducing the number of incorrect diagnoses.

Directions of further research include extending the present method to other fields of image and text such as newspaper articles and Web pages. The addition of other inputs, such as voice recognition, may provide further improvements.

Acknowledgments

The research work was partially supported by the Sayling Wen Cultural & Educational Foundation through the Service Science Research Center (SSRC) at National Chengchi University in Taiwan. I thank J. C. Lin for his assistance in the experiments.

References

[1] Clinico-pathological conference. Johns Hopkins University School of Medicine, 2008. http://oac.med.jhmi.edu/cpc.

[2] K. Barnard and M. Johnson. Word sense disambiguation with pictures. *Artificial Intelligence*, 167:13–30, 2005.

[3] M. Beers, editor. *The Merck Manual of Medical Information*. Merck Research Laboratories, second edition, 2003.

[4] D. Blei and M. Jordan. Modeling annotated data. In *Proceedings of SIGIR'03*, 2003.

[5] P. Duygulu, K. Barnard, J. de Freitas, and D. Forsyth. *Computer Vision - ECCV 2002: 7th European Conference on Computer Vision*, volume 2353, chapter Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary, pages 97–112. Springer-Verlag GmbH, 2002.

[6] R. C. Erdur and I. Seylan. The design of a semantic web compatible content language for agent communication. *Expert Systems*, 25(3):268–294, 2008.

[7] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(5):530–549, May 2004.

[8] J. McCarthy. Notes on formalizing context. *In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 1993.

[9] Y. J. Park, B. C. Kim, and S. H. Chun. New knowledge extraction technique using probability for case-based reasoning: application to medical diagnosis. *Expert Systems*, 23(1):2–20, 2006.

[10] G. Salton and M. McGill, editors. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.

[11] R. Schwitter and M. Tilbrook. Meaningful web annotations for humans and machines using controlled natural language. *Expert Systems*, 25(3):253–267, 2008.

[12] A. Segev, M. Leshno, and M. Zviran. Context recognition using internet as a knowledge base. *Journal of Intelligent Information Systems*, 29(3):305–327, 2007.

[13] A. Segev, M. Leshno, and M. Zviran. Internet as a knowledge base for medical diagnostic assistance. *Expert Systems with Applications*, 33(1):251–255, 2007.

[14] R. Srihari and D. Burhans. Visual semantics: Extracting visual information from text accompanying pictures. In *National Conference on Artificial Intelligence (AAAI-94)*, pages 793–798, 1994.

[15] E. D. Ubeyli. Comparison of different classification algorithms in clinical decision-making. *Expert Systems*, 24(1):17–31, 2007.

[16] R. Valencia-Garcia, J. T. Fernandez-Breis, J. M. Ruiz-Martinez, F. Garcia-Sanchez, and R. Martinez-Bejar. A knowledge acquisition methodology to ontology construction for information retrieval from medical documents. *Expert Systems*, 25(3):314–334, 2008.

[17] S. Wachsmuth and G. Sagerer. Bayesian networks for speech and image integration. In *National Conference on Artificial Intelligence (AAAI-02)*, pages 300–306, 2002.

[18] Wikipedia. Receiver operating characteristic — wikipedia, the free encyclopedia, 2008. [Online; accessed 12-December-2008].