UNIVERSITY OF SOUTHAMPTON

Generating Category-based Documents for Image Queries from Web-based data using their Semantic Representation

by

David R. Hardoon, Sandor Szedmak & John S. Shawe-Taylor

Technical Report

Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science
Image, Speech and Intelligent Systems Group

July 26, 2005

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE
IMAGE, SPEECH AND INTELLIGENT SYSTEMS GROUP

by David R. Hardoon, Sandor Szedmak & John S. Shawe-Taylor

Online images and their surrounding text present a particularly complex problem in image annotation. The surrounding text may only contain partial information about the image and most likely relate to the image in the general context. In this paper we propose an approach to learn the association between images and their surrounding text to automatically generate category-based documents to new image queries. The document generation is done without any image-word annotation before or during the training. We learn a semantic representation between the images and their associated documents using kernel Canonical Correlation Analysis. The semantic space provides a common representation and enables a comparison between the documents and images. This representation is then used for the generation of a new document that best fits the image query. We use text frequency and Term Frequency Inverse Document Frequency as our word representation and compare our proposed method with a standard cross-representation retrieval technique known as Generalised Vector Space Model.

Contents

List of Figures

List of Tables

Acknowledgements

The authors would like to acknowledge the financial support of EU Project LAVA, No. IST-2001-34405.

Chapter 1

Introduction

Due to an increasing rise of multimedia data that is available both on-line and off-line, we are faced with the problematic issue of our ability to access or make use of this information, unless the data is organised in such a way that allows efficient browsing, searching and retrieval. One of these issues is image labelling or multi-labelling where we would like to annotate an image with several keywords that best describe it. A recent solution proposed by (Barnard et al., 2003) is image segmentation with key words associated to the different segmented parts of the image.

In this work we address a different type of problem, where the text associated with the image is not keywords, but a descriptive portion of text. For example news reports using pictures to illustrate a description in the text. Unlike in (Barnard et al., 2003) where the keywords gave a good estimate of the image segments of which they were pre-assigned. The text may not necessarily give a good content description of the image. We select this type of problem as we believe it to be a good representation of the real image-text data available online, which is usually not well captioned or will have a portion of 'general' text that references the image. We define the text associated to an image as a **document** and a document to be comprised of **words**. We propose an extension of (Hardoon and Shawe-Taylor, 2003) whereby using the properties of Kernel Canonical Correlation Analysis (KCCA) to generate new documents to unseen query images. As the original document associated to an image may not be of good description of the image, we suggest evaluating the proposed approach by relevance to mutual categories between the new generated document and image query.

In previous work (Hardoon and Shawe-Taylor, 2003) we presented an approach based on KCCA using the content of both views to retrieve images. This was based on a text query by looking for the highest weighting between the text query and the test images in the feature space. We could reverse the system and look for the best matching document in the given testset to an image query. Although

- One may not have documents that fit the query image, other than those in the training-set.
- We would like to annotate an image query independently from the given set of words within a specific document.

For this purpose we create a new document d^* which contains the words that best fit the query image. It is important to state that during the training stage we do not do any word annotation to the images other than using KCCA to find a common feature representation between the documents, and hence words, to the images. The novelty of the paper is the ability of creating a new document corresponding to an image query neither looking at the training dataset nor using pre-association of words to the images.

The paper is divided as follows, Section 2 gives a brief introduction to CCA and for brevity we exclude the full kernelisation of CCA. In Section 3 we present our proposed method for generating a new document d^* which best fits the query image. In Section 4 we present our experiments and results, while in Section 5 we conclude and discuss future work.

Chapter 2

Canonical Correlation Analysis

Proposed by H. Hotelling in 1936 Canonical correlation analysis can be seen as the problem of finding basis vectors for two sets of variables such that the correlation between the projections of the variables onto these basis vectors are mutually maximised. Correlation analysis is dependent on the co-ordinate system in which the variables are described, so even if there is a very strong linear relationship between two sets of multidimensional variables, depending on the coordinate system used, this relationship might not be visible as a correlation. Canonical correlation analysis seeks a pair of linear transformations one for each of the sets of variables such that when the set of variables are transformed the corresponding coordinates are maximally correlated.

$\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product of the vectors $\mathbf{x}, \mathbf{y}$ and is equal to $\mathbf{x}'\mathbf{y}$. Where A' to denote the transpose of a vector or matrix A. Consider a multivariate random vector of the form $(\mathbf{x}, \mathbf{y})$. Suppose we are given a sample of instances $S = ((\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_l, \mathbf{y}_l))$ of $(\mathbf{x}, \mathbf{y})$, let S_x denote $(\mathbf{x}_1, \ldots, \mathbf{x}_l)$ and similarly S_y denote $(\mathbf{y}_1, \ldots, \mathbf{y}_l)$. We can consider defining a new co-ordinate for $\mathbf{x}$ by choosing a direction $\mathbf{w}_x$ and projecting $\mathbf{x}$ onto that direction $\mathbf{x} \to \langle \mathbf{w}_x, \mathbf{x} \rangle$ if we do the same for $\mathbf{y}$ by choosing a direction $\mathbf{w}_y$ we obtain a sample of the new $\mathbf{x}$ co-ordinate. Let $W_x = \{\mathbf{w}_x^1, \ldots, \mathbf{w}_x^l\}$ and similarly for W_y, where for $k, j = 1 \ldots l$ and $k \neq j$

$$\begin{aligned} \left\langle \mathbf{w}_x^k, \mathbf{w}_x^j \right\rangle &= 0 \\ \left\langle \mathbf{w}_x^k, \mathbf{w}_y^j \right\rangle &= 0 \\ \left\langle \mathbf{w}_y^k, \mathbf{w}_y^j \right\rangle &= 0 \end{aligned}$$

Let $S_{x,W_x} = \left(\langle \mathbf{w}_x^1, \mathbf{x}_1 \rangle, \ldots, \langle \mathbf{w}_x^l, \mathbf{x}_l \rangle\right)$ with the corresponding values of the new $\mathbf{y}$ co-ordinate being $S_{y,W_y} = \left(\langle \mathbf{w}_y^1, \mathbf{y}_1 \rangle, \ldots, \langle \mathbf{w}_y^l, \mathbf{y}_l \rangle\right)$. The function to be maximised for a

single direction is (we drop the upper script for clarity)

$$\rho = \max_{\mathbf{w}_x, \mathbf{w}_y} corr(S_{x,\mathbf{w}_x}, S_{y,\mathbf{w}_y}) = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\langle S_{x,\mathbf{w}_x}, S_{y,\mathbf{w}_y} \rangle}{\|S_{x,\mathbf{w}_x}\| \|S_{y,\mathbf{w}_y}\|}.$$

For brevity we solve the problem for a single direction, although in practice one is able to compute all directions. We use $\hat{\mathbb{E}}\left[f(\mathbf{x}, \mathbf{y})\right]$ to denote the empirical expectation of the function $f(\mathbf{x}, \mathbf{y})$, where $\hat{\mathbb{E}}\left[f(\mathbf{x}, \mathbf{y})\right] = \frac{1}{l} \sum_{i=1}^{l} f(\mathbf{x}_i, \mathbf{y}_i)$. We can rewrite the correlation expression as

$$\begin{aligned}
\rho &= \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\hat{\mathbb{E}}[\langle \mathbf{w}_x, \mathbf{x} \rangle \langle \mathbf{w}_y, \mathbf{y} \rangle]}{\sqrt{\hat{\mathbb{E}}[\langle \mathbf{w}_x, \mathbf{x} \rangle^2] \hat{\mathbb{E}}[\langle \mathbf{w}_y, \mathbf{y} \rangle^2]}} \\
&= \frac{\hat{\mathbb{E}}[\mathbf{w}_x' \mathbf{x} \mathbf{y}' \mathbf{w}_y]}{\sqrt{\hat{\mathbb{E}}[\mathbf{w}_x' \mathbf{x} \mathbf{x}' \mathbf{w}_x] \hat{\mathbb{E}}[\mathbf{w}_y' \mathbf{y} \mathbf{y}' \mathbf{w}_y]}} \\
&= \frac{\mathbf{w}_x' C_{\mathbf{xy}} \mathbf{w}_y}{\sqrt{\mathbf{w}_x' C_{\mathbf{xx}} \mathbf{w}_x \mathbf{w}_y' C_{\mathbf{yy}} \mathbf{w}_y}}.
\end{aligned}$$

The total covariance matrix is a block matrix where the within-sets covariance matrices are $\mathbf{C_{xx}}$ and $\mathbf{C_{yy}}$ and the between-sets covariance matrices are $\mathbf{C_{xy}} = \mathbf{C'_{yx}}$, although this is subjected to the data having a zero-mean.

The dual form of CCA can be formulated by expressing the image weights W_x and document weights W_y as a linear combination of the training examples $W_x = X\alpha$ and $W_y = Y\beta$ where X and Y are matrices with rows $\{\mathbf{x_i}, \ldots, \mathbf{x_l}\}$ and $\{\mathbf{y_i}, \ldots, \mathbf{y_l}\}$ respectively. Let $W_x = \left\{\mathbf{w_x^1}, \ldots, \mathbf{w_x^l}\right\}$ and $W_y = \left\{\mathbf{w_y^1}, \ldots, \mathbf{w_y^l}\right\}$ as we obtain a set of weight vectors for each sample. Respectively α and β are a sequence of feature vectors such that $\alpha = \{\boldsymbol{\alpha^1}, \ldots, \boldsymbol{\alpha^l}\}$ and $\beta = \{\boldsymbol{\beta^1}, \ldots, \boldsymbol{\beta^l}\}$.

Following (Hardoon et al., 2004; Shawe-Taylor and Cristianini, 2004, for full derivation) the dual form of CCA with regularisation parameter τ will be given by solving

$$\max_{\alpha,\beta} \rho = \frac{\alpha' K_x K_y \beta}{\sqrt{((1-\tau)\alpha' K_x^2 \alpha + \tau \alpha' K_x \alpha)((1-\tau)\beta' K_y^2 \beta + \tau \beta' K_y \beta)}}.$$

Where K_x is the kernel matrix for the images and K_y the respective kernel matrix for the documents.

Chapter 3

Generating a New Document

We are faced with the problem of creating a new document d^* that best matches our image query. Based on the idea of CCA we are looking for a vector that has maximum covariance to the query image with respect to the weight matrices α and β. Let $f = K_x^i \alpha$, where the vector K_x^i contains the kernelised inner products between the query image i and the images occurring in the training set. We have

$$\max_{d^*} \left\langle f, W_y' d^* \right\rangle,$$

where W_y is as defined in the previous section.

Assume that the new document is represented in the form $d^* = D\hat{d}$, where D is a design matrix of size $n \times m$, where n is the number of known words in the training dataset and m is an arbitrary number depending on the expected structure of the document. For simplicity we look at the case where the expected structure of the document is a of a single word which is the most relevant word for the query image. Therefore the structure of our design matrix is the identity matrix $D = I$ of size $n \times n$. We may say that the vector $\hat{d}$ gives a convex combination of the columns of the identity matrix, thus it satisfies the constraints

$$\sum_{i=1}^{n} \hat{d}_i = 1, \quad \hat{d}_i \geq 0 \quad i = 1, \ldots, n. \tag{3.1}$$

The problem becomes

$$\max_{\hat{d}} f' W_y' \hat{d}$$

under the same constraints. Let $c = f' W_y'$ we have

$$\max_{\hat{d}} c\hat{d}.$$

Due to the constraints in equation (5) the components of the optimum solution d^* are equal to

$$(d)_i^* = \begin{cases} 1 & i = \arg\max_j c_j, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

This generates a document containing a single word. We modify the original maximisation problem to relax the optimum solution to include words above a threshold T

$$(d)_i^* = \begin{cases} 1 & i = (c_j \geq T), \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

The new relaxed formulation will generate a document with varying number of words, depending on T, we are able to use c_j to rank the relevance of the words. Equation (3.3) is our chosen approach for generating documents for the experiments in the following section.

If the design matrix D is the identity matrix then we receive a document comprising of one word only, as shown in equation (3.2). We propose an approach for creating a document more then a single word by modifying the design matrix. Let l be a given number, chosen arbitrarily, if every column of D contains 1 in l components and all other components are 0 and all possible but different vectors are included with this property in D then one can derive the best fitting document containing l words. It is easy to show that the optimal solution for this design matrix gives the column of D corresponding to the l greatest components of the vector $f'W_y$, hence the l best words fitting to the query image can be found by a single sorting procedure.

Chapter 4

Experiments

In the following experiments the problem of learning the semantics of multimedia content by combining image and text data is addressed. The learnt semantics is then applied to generate documents to query images. The aim is to allow the retrieval of words from an image query without reference to any original labelling associated with the image. We use a combined multimedia image-text web database which manifests the introduced problem. The database was kindly provided by the authors of (Kolenda et al., 2002). The data is divided into three classes: Sport, Aviation and Paintball, 400 records each and consisted of jpeg images retrieved from the Internet with attached text, with an overall of 1200 samples. We randomly split each class into two halves, obtaining 600 samples for training and 600 for testing. The extracted features of the images were the HSV (Hue Saturation Values) colour and the image Gabor texture. As discussed in the introduction and shown in figures 4.2-4.6, the words annotated to the images are not a keyword association but more of a general descriptive nature to the image. Therefore our representation of the words is crucial, as we wish to capture the information that relates the words within the documents to the images, we compare two approaches of word representation; The term frequency vector which is the number of occurrences of the word in the document, and Term Frequency Inverse Document Frequency (TFIDF) (Salton and McGill, 1983) which is

$$\begin{aligned} \text{TFIDF}(d_i, w_j) &= (\text{number of } w_j \text{ in } d_i) \\ &\quad \times \log\left(\frac{N}{\text{number of documents that contain } w_j}\right), \end{aligned}$$

where N is the number of documents, d_i is document i and w_j is word j.

We compare the performance of our method with a retrieval technique based on the Generalised Vector Space Model (GVSM)(Wong et al., 1985). This uses as a semantic feature vector, the vector of inner products between either a query image and each training image or test documents and each training document. For both KCCA and

TABLE 4.1: Confusion matrix of the # of words in each subset.

	A	B	C
A	1269	447	324
B	447	850	86
C	324	86	327

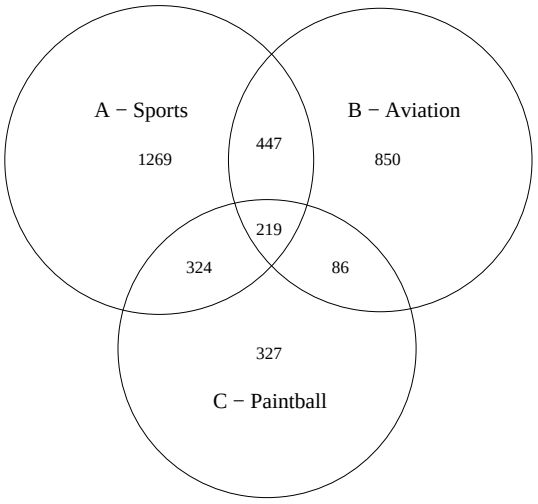

FIGURE 4.1: **We separate the dictionary into its categories with overlapping regions.**

GVSM the first view was obtained by the combination of a Gaussian kernel (with σ as the minimum distance between the different images) with a linear kernel on the Gabor textures, while the second view was obtained by a linear kernel on the term frequencies or TFIDF features. We compute the KCCA regularisation parameter τ as described in (Hardoon and Shawe-Taylor, 2003).

Let $\mathcal{W}$ be the set of all words in our database comprising an overall of 3522 words and let $A, B, C \subseteq \mathcal{W}$ be such that A is the Sports category with a total of 2259 words, B the Aviation category with a total of 1602 words and C the Paintball category with a total of 956 words. As shown in table 4.1 and figure 4.1 some of the words overlap categories, although there are only 219 words overall which are common to all three categories. The words in $\mathcal{W}$ are associated with the categories by the following approach; for example, if word w is associated with an image that belongs to category A, w will belong to A as well. Table 4.2 shows an example of the different possible words in the categories, as visible some words such as "accounted" in the Sports category would not have been placed there via human selection. The word-image association is limited to the contents of the website.

After generating the new document d^*, as described in Section 3, we try and evaluate the relevance of the words within the document to the image query. As we attempt

TABLE 4.2: Example of words in categories:

Sports	**aa, aaron, abdominal, abdul, abdur, ability, abroad, absence, accomplish, accounted, achieve, achilles ...**
Aviation	**air, airborne, aircraft, airesearch, airflow, airframe, airliner, airmotive, airpark, areonautical, airline ...**
Paintball	**warriors, watch, weather, web, wing, wednesday, white, wildcards,wildfire, winning, wizard, wordogz, world ...**

an automatic word retrieval where human intervention is at a minimum the definition of a "relevant" word is not trivial. We attempt this definition by denoting three levels of relevance, the first - level 1; If a word is in a category (including overlapping ones) and is the same category as the image, it is considered a success with a weighting of 1 otherwise the word will be considered a mistake and will be of weighting 0. In level 2 we follow the weightings scheme of level 1 but start penalising the words that are in the overlapping categories by -0.25 such that if a word belongs to either of the two overlapping categories it will be given a weighting of 0.75 and a weighting of 0.5 if it belongs to all three. In the last and final level, level 3, we further increase the penalty to -0.5 such that if a word belongs to either of the two overlapping categories it will have a weighting 0.5 and if the word belongs to all three it will be considered a mistake (a weighting of 0). The choice of the penalty is arbitrary. The success of the category-based document generation is computed as the sum of the weights of the words in the documents averaged over the number of words in the documents.

The experiments, in both methods, were repeated 10 times and averaged over all the query test images. In each repeat we evenly and randomly split the testing and training samples. As shown in Hardoon and Shawe-Taylor (2003) we do not need to test for all the eigenvectors as the success rate will generally saturate after half the number of eigenvectors. Therefore we suffice in testing for a selection of $1 - 300$ eigenvectors and picking the best one.

We expect that for some of the image queries the words within the generated documents would overlap. Although, as we would like to show that our approach does not generate the trivial solution (i.e. the same words within the documents for all queries). We compute the variance, by variance we mean the difference between words in each generated document, of the correctly retrieved words within a document. Let

$$\text{variance} = \frac{\text{number of different correct words found}}{\text{all correct words found}}.$$

In the case of retrieving 10 words within a document we are unable to avoid overlapping words, therefore we normalise the variance by the maximum different words possible, which amounts to - $(\frac{3522}{600 \cdot 10})$.

In tables 4.3 and 4.5 we present the KCCA comparison between the document generation of 1 and 10 words. Table 4.3 presents the results obtained by using term frequency, while table 4.5 presents those obtained by using TFIDF. The best success rates for the specific task are highlighted. We arbitrarily choose the eigenvector selection that gives the best retrieval result. We find that the TFIDF representation of the documents outperforms that of the term frequency representation, except for level 1 where term frequency obtains a higher level of retrieval with a lower level of variance. In tables 4.3 and 4.5 'eig' are the corresponding eigenvectors used in the feature projection.

In tables 4.4 and 4.6 we present the baseline approach using term frequency and TFIDF. We are able to observe that with GSVM using term frequency obtains a higher level of success while TFIDF obtains a higher level of variance. Although a high success rate in level 1 using term frequency representation we find that the variance of the words are extremely low suggesting that GSVM generates documents with the same words for most of the queries. Clearly, KCCA is consistently better than the baseline method in both approaches with relatively good results.

We provide further analysis of the KCCA results. Although the TFIDF features need a larger number of eigenvectors for the feature projection it is able to produce a higher success rate then that of the term frequency vector, except in level 1 although the low variance implies that the retrieved words are very similar. We see that as we increase the weight of the penalty TFIDF is able to generate documents with words that are more singular to the topic and are of a higher variance. In previous work Hardoon and Shawe-Taylor (2003) we have shown that increasing the number of eigenvectors for the feature selection will increase the success rates of the content-based image retrieval task, as visible in Tables 4.3 and 4.5 we can see a similar effect on the variance of the retrieved keywords.

Figures 4.2, 4.3 , 4.4 and 4.5 show examples of query images with their original text and the generated documents with 10 words, using KCCA level 1 weighting scheme with TFIDF. The words that belong to the same category as the image are in bold while those which are mistakes are italicised. Figure 4.6 shows an example of an image with a very complicated text assigned to it. Several of the shown figures present the problem that the original text may not be informative to the image, this also illustrates why we do not asses the number of words in-common with the original text. We are not trying to recreate the original query image text but generate a new most relevant one according to the learnt semantic space.

TABLE 4.3: KCCA: Success results using term frequency. (eig - eigenvectors)

	Documents with 1 Word		
	Success	Variance	# eig
Level 1	**96.27**%	0.77%	3
Level 2	73.17%	31.17%	270
Level 3	51.19%	37.2%	273
	Documents with 10 words		
	Success	Variance	# eig
Level 1	**89.51**%	1.65%	5
Level 2	71.77%	21%	283
Level 3	45.45%	2.81%	6

TABLE 4.4: GSVM: Success results using Term Frequency

	Documents, 1 word		Documents , 10 words	
	Success	Variance	Success	Variance
Level 1	80.07%	0.41%	61.69%	0.58%
Level 2	52.43%	0.63%	44.55%	0.8%
Level 3	24.79%	0.68%	27.4%	1.03%

TABLE 4.5: KCCA: Success results using TFIDF features.(eig - eigenvectors)

	Documents with 1 word		
	Success	Variance	# eig
Level 1	88.14%	32.55%	264
Level 2	**75.69**%	38.93%	278
Level 3	**63.34**%	41.75%	278
	Documents with 10 words		
	Success	Variance	# eig
Level 1	86.22%	20.02%	299
Level 2	**72.75**%	23.72%	299
Level 3	**59.24**%	25.74%	296

TABLE 4.6: GSVM: Success results using TFIDF

	Documents, 1 word		Documents, 10 words	
	Success	Variance	Success	Variance
Level 1	35.62%	0.95%	54.33%	0.87%
Level 2	35.62%	0.95%	43.21%	1.1%
Level 3	35.62%	0.95%	32.09%	1.15%

FIGURE 4.2: Aviation: Original text "**is posed as if its nose gear has collapsed. executive decision is to the right of center. the 747-200 that appeared in the rookie appears at center. the convair 880 that was used in speed**" generated document with 10 words, from highest to lowest rank "**convair, museum, cccp, eagle, voyage, cv, protivophozharny, tower, pima, roll**".

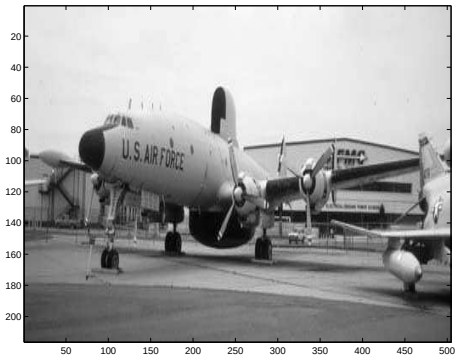

FIGURE 4.3: Aviation: Original text "**ec-121k, 141309, c/n 4433 . air force museum's page about ec-121d, 53-0555 ec-121k, 141309 at the mcclellan afb museum on april 3, 1993. it was built as a navy wv-2, but it is displayed as air force ec-121d, 53-0552. its lockheed construction number is 4433.**" generated document with 10 words, from highest to lowest rank "**museum, commando, goodyear, pima, page, takes, link, air, afb, castle**".

FIGURE 4.4: Paintball: Original text "**benini reffing**" generated document with 10 words, from highest to lowest rank "**fate, darkside, kc, strange, team, wildcards, takeover, avljalde, hostile, check**".

FIGURE 4.5: Paintball: Original text "**all americans**" generated document with 10 words, from highest to lowest rank "**ref, farside, american, team, trauma, stay, leader, flag, takeover, avljalde**".

FIGURE 4.6: Sports: Original text "**ap photo more photos february 18, 2002 toronto (ap) – sam cassell 's injured toe didn't hurt his effectiveness against the toronto raptors . but he might be selective about future games he plays in so he's ready for the playoffs. "i'd rather take care of it now," said cassell, who missed two games with a sprained left big toe before scoring 20 points in the milwaukee bucks ' 91-86 victory over the toronto raptors on sunday. "this is not a joke," he said. "this is probably the worst i've felt as a professional basketball player. it's painful every step you take. coach says, 'you can take the pain!' but not this kind of pain." cassell, who scored eight points in the fourth quarter, said he would need 20 days to heal. "we have a game (monday)** ..." generated document with 10 words, from highest to lowest rank "*boyz*, **attitude, pt, hot,** *urban, quest ,rip,* **team,ap,***matrix*".

Chapter 5

Conclusions

The problem of retrieving information via content is a non trivial one. We present a relatively simple technique to generate documents to an image query with neither reference to the training documents nor the usage of keyword assignment to the image or to image segments. In the presented work we differ from the conventional image-word retrieval problem, which aims to associated a word which best describe the actual content of the image. We are interested in associating the content of a genre to an image. This can be used to generate documents (information) to an image based on the category to which they are presented to. For example, an image of Jerusalem will generate documents relevant to the Middle East conflict if fed to a News website, while if fed to a travel agency will most likely generated documents relevant to travel in and around Jerusalem, etc.

We learn an association by using kernel Canonical Correlation Analysis to find a semantic representation which is common to both views. We find that although the simplicity of the approach, our results are promising and better than those obtained by the baseline method. Although several issues remain, such as; how one can define a better "relevance" test for the retrieved keywords? As we may also have image queries that do not have their associated words in the training corpus. It may also be relevant to devise a probabilistic scheme for the word penalty rather then using an arbitrary one. A further avenue that could be looked at, is the method of creating the new document d^*. Usage of better image features would be investigated in future work.

With these open issues we believe that this is an interesting problem addressed from an unconventional perspective, as we only relay on the given information and attempt to infer from it. We have used a difficult database which we believe to manifest a real-world scenario that is encountered daily on the Internet. We would like to apply it to a more generic type of database such as News. Although, it also would be interesting to observe the performance of this proposed system an artificial database where images are annotated with keyword captions that give a good concise description of the image.

Bibliography

Shotaro Akaho. A kernel method for canonical correlation analysis. In *International Meeting of Psychometric Society*, Osaka, 2001.

Francis Bach and Michael Jordan. Kernel independent component analysis. *Journal of Machine Leaning Research*, 3:1–48, 2002.

Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Fretias, David M. Beli, and Michael I. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.

Magnus Borga. *Learning Multidimensional Signal Processing*. PhD thesis, Linkping Studies in Science and Technology, 1998.

Colin Fyfe and Pei Ling Lai. Ica using kernel canonical correlation analysis. *Proceedings of International Workshops on Independent Component Analysis and Blind Signal Separation*, pages 279–284, 2000.

Colin Fyfe and Pei Ling Lai. Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems*, 2001.

A. Gifi. *Nonlinear Multivariate Analysis*. Wiley, 1990.

David R. Hardoon and John Shawe-Taylor. KCCA for different level precision in content-based image retrieval. In *Proceedings of Third International Workshop on Content-Based Multimedia Indexing*, IRISA, Rennes, France, 2003.

David R. Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: an overview with application to learning methods. *Neural Computation*, 16: 2639–2664, 2004.

T. Kolenda, L. K. Hansen, J. Larsen, and O. Winther. Independent component analysis for understanding multimedia content. In H. Bourlard, T. Adali, S. Bengio, J. Larsen, and S. Douglas, editors, *Proceedings of IEEE Workshop on Neural Networks for Signal Processing XII*, pages 757–766, Piscataway, New Jersey, 2002. IEEE Press. Martigny, Valais, Switzerland, Sept. 4-6, 2002.

G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Berlin, 1983.

John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

S. Wong, W. Ziarko, and P. Won. Generalized vector space model in information retrieval. *Proceedings of the 8th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 11:18–25, 1985.