Chapter 6
Fusion Techniques for Combining Textual and Visual Information Retrieval

Adrien Depeursinge and Henning Müller

Abstract This chapter describes several approaches for information fusion that have been used in ImageCLEF over the past seven years. In this context, the fusion of information is mainly meant to combine textual and visual retrieval. Data fusion techniques from 116 papers (62% of ImageCLEF working notes) are categorized, described and discussed. It was observed that three general approaches were used for retrieval that can be categorized based on the system level chosen for combining modalities: 1) at the input of the system with inter–media query expansion, 2) internally to the system with early fusion and 3) at the output of the system with late fusion which is by far the most widely used fusion strategy.

6.1 Introduction

Any concept with even a low level of semantics is best described by the co–occurrence of several events in multiple sources of information. In medicine for instance, diagnosis is established with confidence if, and only if, the laboratory results, the history of the patient and possibly radiographic examinations are all taken into account and converge to a unique conclusion. In another context, a photograph of a football game can be associated with its corresponding event only when the date and the place are known. Consequently, computerized Information Retrieval (IR) must be able to fuse multiple modalities in order to reach satisfactory performance. Information fusion has the potential of improving retrieval performance by relying on the assumption that the heterogeneity of multiple information sources

Adrien Depeursinge
University and University Hospitals of Geneva (HUG), Rue Gabrielle–Perret–Gentil 4, 1211 Geneva 14, Switzerland, e-mail: adrien.depeursinge@sim.hcuge.ch

Henning Müller
Business Information Systems, University of Applied Sciences Western Switzerland (HES–SO), TechnoArk 3, 3960 Sierre, Switzerland, e-mail: henning.mueller@hevs.ch

H. Müller et al. (eds.), *ImageCLEF*, The Information Retrieval Series 32, 95
DOI 10.1007/978-3-642-15181-1_6,

and/or algorithms allow cross–correction of some of the errors, leading to better results. Multiple views of the problem potentially allow a reduction of the semantic gap, which is defined in image retrieval as the discrepancy between the user's intentions when searching for a particular image and the visual information that the features are able to model (Smeulders et al, 2000).

Multi–modal information is often available in digital repositories. For example, videos are constituted by synchronized visual and audio modalities. Frequently, images on the Internet come with textual annotations that are semantically related. Modern health information systems enable access to structured information (e.g. age of the patient, gender, laboratory results), free–text in reports, radiological images and biosignals such as electrocardiograms. This means that the major challenge in information fusion is to find adapted techniques for federating multiple sources of information for either decision–making or information retrieval. Fusing multiple information sources is not devoid of risks. Two aspects require particular attention when performing information fusion in order to avoid degradation of the system performance:

- the relevance of all modalities to be fused must be verified to prevent the introduction of noise into the system;
- the fusion scheme must be able to assess trustworthiness of the modalities towards the query in order to allocate confidence in modalities that have high relevance in the context of the query.

Information fusion has been a lively research topic during the last 20 years (see, e.g. (Saracevic and Kantor, 1988; Belkin et al, 1993, 1994; Shaw and Fox, 1994)). Fusion was carried out at three different levels of an IR system (Frank Hsu and Taksa, 2005):

- at the input of the IR system while using multiple queries or query expansion;
- within the system where several algorithms and/or features can be used to increase the heterogeneity of results (i.e. boosting or multiple classifier systems);
- at the output of the system when combining several lists of documents.

Investigation of the effectiveness of combining text and images for retrieval including medical image retrieval is one of the main goals of the ImageCLEF campaign (Hersh et al (2007)). Since its first year in 2003, the organizers of ImageCLEF provided multimedia databases containing images with associated text thus allowing for multi–modal retrieval. During the past seven years of ImageCLEF, three image retrieval tasks elicited research contributions in fusion techniques for combining textual and visual information retrieval:

- the photo retrieval task proposed since 2003,
- the medical image retrieval task proposed since 2004,
- Wikipedia image retrieval task proposed since 2008.

In total, 116 (62%) out of 187 papers in ImageCLEF submissions from 2003 to 2009 attempted to mix Text–Based Image Retrieval (TBIR) with Content–Based Image Retrieval (CBIR) to investigate the complementarity of the two modalities (see Table 6.1).

Table 6.1: Number of papers per task and per year merging textual and visual information during the past seven years of ImageCLEF.

	2003	2004	2005	2006	2007	2008	2009
photo	0/4 (0%)	6/12 (50%)	6/11 (54%)	4/12 (33%)	14/19 (74%)	18/25 (72%)	12/16 (75%)
medical	–	6/11 (54%)	10/14 (71%)	8/10 (80%)	7/9 (78%)	6/11 (54%)	8/14 (57%)
Wikipedia	–	–	–	–	–	8/11 (73%)	3/8 (38%)

6.1.1 Information Fusion and Orthogonality

From a certain point of view, all systems that are using more than one single feature are carrying out information fusion. However, features within a modality may be strongly correlated among them (e.g. consecutive bins of a color histogram, see Depeursinge et al (2010–to appear)). As a consequence, the rank of the space spanned by the feature vector $\mathbf{v_A} = \{a_1 \dots a_{N_A}\}$ of the modality A is usually much inferior the number of feature N_a of A. We have:

$$rank(A) \ll N_a. \tag{6.1}$$

While taking into account M modalities $\{A_1 \dots A_M\}$ defined by their respective feature vectors $\{\mathbf{v_{A_1}} \dots \mathbf{v_{A_M}}\}$, the linear dependence of multi–modal space is given by the number L of possible solutions $(x_1, x_2, \dots, x_M)$ over all realizations of $\{\mathbf{v_{A_1}} \dots \mathbf{v_{A_M}}\}$:

$$x_1\mathbf{v_{A_1}} + x_2\mathbf{v_{A_2}} + \cdots + x_M\mathbf{v_{A_M}} = \mathbf{0}, \tag{6.2}$$

with $x_1, x_2, \dots, x_M \in \mathbb{R} \setminus 0$. Thereby, the amount of heterogeneity H of a combination of modalities can be measured using the number P of linearly independent vectors divided by the number of modalities M:

$$H = \frac{P}{M}. \tag{6.3}$$

H has values in $[0;1] \setminus 0$ and can be seen as the inverse of redundancy. It is important to note that large values of H would not be desirable as it means that no redundancy occurs in the set of modalities, which means that at least M-1 modalities are not related to any concept (or class). An ideal multi–modal system should be composed of modalities that are correlated for no other reason than that these are all related to a corpus of concepts. This was observed by Lee (1997) who stated that "different modalities might retrieve similar sets of relevant documents but retrieve different sets of non–relevant documents". This means that the information gain I_G (according to Quinlan (1986)) of the features from each modality towards the corpus of concepts must be above a critical threshold. I_G was originally defined by Quinlan to iteratively choose informative attributes to build decision trees. $I_G(Y|X)$ of a given attribute X with respect to the class attribute Y quantifies the change in information entropy when the value of X is revealed:

$$I_G(Y|X) = H(Y) - H(Y|X). \tag{6.4}$$

The information entropy $H(Y)$ measures the uncertainty about the value of Y and the conditional information entropy $H(Y|X)$ measures the uncertainty about the value of Y when the value of X is known:

$$H(Y) = -\sum_{y\in\mathscr{Y}} p(y)\log p(y), \tag{6.5}$$

$$H(Y|X) = -\sum_{x\in\mathscr{X},y\in\mathscr{Y}} p(x,y)\log p(y|x). \tag{6.6}$$

To summarize, an optimal multi–modal system should maximize the degree of heterogeneity H while maximizing the information gain I_G of each modality (taken independently) towards the studied corpus of classes.

6.2 Methods

The techniques used through the seven past years in ImageCLEF for fusing textual and visual image information were reviewed and categorized based on their similarities. Only papers that mixed textual and visual retrieval were studied and papers using multiple classifier systems on one single modality were left aside.

In total, techniques from 116 papers from 2004 to 2009 were categorized in the subsections of Section 6.3. An overview of the techniques and trends is presented. Justifications for the approaches and generally known problems are discussed in Section 6.4.

6.3 Results

The various techniques used for fusing textual and visual information in ImageCLEF are described in this section. When available, comparisons of the performances among techniques are detailed. A global view of the data fusion techniques is proposed in Section 6.3.5.

6.3.1 Early Fusion Approaches

An early fusion consists of mixing modalities before making any decisions. The combination takes place in the feature space where the textual and visual attributes

($\{t_1 \ldots t_k\}$ and $\{v_1 \ldots v_l\}$ respectively) are concatenated into one vector to create one unique feature space $\{t_1 \ldots t_k \quad v_1 \ldots v_l\}$ (see, e.g. (Snoek et al, 2005; Gunes and Piccardi, 2005; Depeursinge et al, 2010–to appear)). It enables a true multimedia representation where one decision rule is based on all information sources. The major drawback of this method is that it is confronted with the curse of dimensionality as the the dimension of the resulting feature space is equal to the sum of the dimensions of the subspaces **t** and **v**. High–dimensional spaces tend to scatter the homogeneous clusters of instances belonging to the same concepts. This has to be handled using an appropriate feature weighting scheme, which is usually difficult to achieve in practice for complex multi-class problems where the majority of features are important to predict one particular class but introduce noise for all the other classes.

Early fusion is used without any feature weighting in Ferecatu and Sahbi (2008) in the photo retrieval task where text and visual features are simply normalized before being concatenated. A comparison with a late fusion method based on the combMIN rule (see Section 6.3.2.6) shows that the early fusion performs slightly better but without statistical significance.

Early fusion using various feature weighting schemes for medical image retrieval is investigated in (van Zaanen and de Croon, 2004; Deselaers et al, 2005; Cheng et al, 2005; Deselaers et al, 2006, 2007). Entropy–based feature weighting methods showed to outperform significantly performance obtained using a single modality in (Deselaers et al, 2006, 2007), which is in accordance with our assumptions in Section 6.1.1 as the information gain I_G is based on entropy measures (see Eq. 6.4).

A degradation of the retrieval performance is observed with the Wikipedia task in (Moulin et al, 2008) where a visual vocabulary is first created from basic image features, which is then fused with text features using a TF–IDF weighting (see (Salton and Buckley, 1988)).

In 2009, the best automatic mixed run of the medical task was based on early fusion of text features with very basic image features modeling color information of the whole image (Berber and Alpkoçak, 2009).

6.3.2 *Late Fusion Approaches*

Late fusion approaches concern every technique for combining outputs of distinct systems. The diversity among late fusion strategies is much broader than the early fusion approach and many techniques for combining lists of documents (runs) were used in ImageCLEF and are detailed in this section.

6.3.2.1 Rank–based Fusion vs. Score–based Fusion

When combining runs from different systems there are two main approaches. The relevance of a document d can be measured by either its rank $R_j(d)$ in the list $L_j(d)$

given by an IR system j or by its score $S_j(d)$ (or relevance, similarity, distance to the query). The score–based strategies, although more common, require a normalization among all systems in order to balance the importance of each of them, which is not the case of the rank–based strategies.

Several approaches are found in the literature for normalizing scores. A commonly used technique called MinMax was proposed by Lee (1997, 1995) where the normalized score $\overline{S}$ is computed as follows:

$$\overline{S} = \frac{S - S_{min}}{S_{max} - S_{min}}, \tag{6.7}$$

with S_{min} and S_{max} the lowest and highest scores found among all runs, systems or topics. Montague and Aslam (2001) also proposed two linear transformations for the normalization of scores: Sum and zero–mean and unit–variance ZMUV. Sum maps S_{min} to 0 and the sum of all scores to 1. In ZMUV, the average of all scores is mapped to 0 and their variance to 1. Sum and ZMUV are mostly intended to be used with the combination techniques combSUM and combMNZ respectively (see Sections 6.3.2.4 and 6.3.2.5).

6.3.2.2 Intersection of Runs

The most straightforward combination rule for multiple runs L_j is to intersect each other. The four combination operators used in ImageCLEF are defined as follows (see Villena-Román et al (2007b,a)):

$$\text{OR} \qquad L_1 \cup L_2, \tag{6.8}$$
$$\text{AND} \qquad L_1 \cap L_2, \tag{6.9}$$
$$\text{LEFT} \qquad (L_1 \cup L_2) \cup (L_1 \setminus L_2), \tag{6.10}$$
$$\text{RIGHT} \qquad (L_1 \cup L_2) \cup (L_2 \setminus L_1). \tag{6.11}$$

Usually these combination operators were associated with reordering rules (see Sections 6.3.2.3, 6.3.2.4, 6.3.2.5, 6.3.2.6 and 6.3.2.7). In Müller et al (2005), the union of runs (OR) is performed by adding various percentages of top textually– and visually–retrieved documents.

6.3.2.3 Reordering

When documents of various lists are gathered, a rule for reordering the documents is required to obtain a final ranking. In (Hoi et al, 2005; Florea et al, 2006; Gobeill et al, 2006; Fakeri-Tabrizi et al, 2008; Simpson et al, 2009; Mulhem et al, 2009; Besançon and Millet, 2005; Zhou et al, 2008a), the textually–retrieved documents are reordered based on their visual score. Inversely, visually–retrieved documents are reordered with their corresponding textual scores in (Villena-Román et al, 2005;

Gobeill et al, 2006; Clinchant et al, 2007; Chang and Chen, 2007; Jensen and Hersh, 2005; Daumke et al, 2006; Hersh et al, 2006; Granados et al, 2008; Ah-Pine et al, 2008, 2009). In Hare et al (2009); Gao and Lim (2009), a text run is reordered to maximize content–based distance among top images to favor the diversity of top–retrieved images.

6.3.2.4 Linear Combinations

In order to reorder documents based on both textual and visual scores S_t and S_v, a commonly used technique for obtaining the final score $S_{mixed}(d)$ of the document d is to perform a linear combination of scores as follows:

$$S_{mixed}(d) = \alpha S_t(d) + (1-\alpha)S_v(d), \tag{6.12}$$

where S_t and S_v are usually normalized and $\alpha \in [0;1]$. Linear combination of scores was used as defined by Equation 6.12 in a large number of papers (37% of the papers dealing with information fusion in ImageCLEF, (Cheng et al, 2004b,a; Müller et al, 2004; Alvarez et al, 2004; Besançon et al, 2004; Lin et al, 2004; Lim and Chevallet, 2005; Chang et al, 2005; Müller et al, 2005; Adriani and Framadhan, 2005; Ruiz and Southwick, 2005; Besançon and Millet, 2005; Díaz-Galiano et al, 2006; Rahman et al, 2006; Lacoste et al, 2006; Gobeill et al, 2006; Wilhelm and Eibl, 2006; Wilhelm et al, 2007; Maillot et al, 2006; Villena-Román et al, 2007b,a; Clinchant et al, 2007; Jair Escalante et al, 2007; Gao et al, 2007; Díaz-Galiano et al, 2007; Zhou et al, 2007; Hoi, 2007; Kalpathy-Cramer and Hersh, 2007; Yamauchi et al, 2008; Zhou et al, 2008a; Díaz-Galiano et al, 2008; Zhou et al, 2008b; Zhao and Glotin, 2008; Navarro et al, 2008c,b; O'Hare et al, 2008; Ah-Pine et al, 2008; Torjmen et al, 2008; Navarro et al, 2008d,a; Rácz et al, 2008; Ye et al, 2009; Ruiz, 2009; Torjmen et al, 2009; Boutsis and Kalamboukis, 2009; Daróczy et al, 2009; Mulhem et al, 2009; Zhou et al, 2009; Jair Escalante et al, 2009)).

Most often, arbitrary values are used for the weight α with usually more weight on textual scores as textual retrieval performs better than content–based retrieval, at least in terms of recall whereas CBIR tends to have higher early precision (see Müller et al (2008); Belkin et al (1994); Shaw and Fox (1994)). An exception was observed by Douze et al (2009) who obtained best results when applying a strong weight for the visual score.

Some groups used data from the previous year to learn weights (Ruiz, 2009). Järvelin et al (2007) computed the weights based on the variation of the modality towards the corpus of classes. In Rahman et al (2007), the weights are updated dynamically based on the user's relevance feedback. Document–specific weighting is used in Granados et al (2008, 2009) where weight of a document in the 'support' modality is divided by its rank.

In order to foster the modality with higher confidence, a linear combination of the scores is used only if both scores S_t and S_v are above a given threshold in Mulhem (2008); Broda et al (2009). The score of only one of the modalities is used otherwise.

In Zuccon et al (2009), text runs are reordered with a linear combination of text score and visual score based on factor analysis and bi–clustering to favor diversity among the retrieved images.

Linear combinations of ranks are much less frequently used, and were tried by Magalhães et al (2007); Jair Escalante et al (2008). Arithmetic and harmonic means of ranks are employed in Glotin and Zhao (2008). Linear combinations based on ranks have the advantage of not requiring a prior normalization. However, the assessment of confidence of the modalities is lost as two images having the same rank in both textual and visual modalities can have very different relevance towards the query.

CombSUM

A particular case of the linear combination is the combSUM rule where the scores of each modality j are summed to obtain the final score:

$$S_{mixed}(d) = \sum_{j=1}^{N_j} S_j(d), \tag{6.13}$$

with N_j the number of modalities to be combined. CombSUM is equivalent to a linear comb with $\alpha = 0.5$ if the scores are normalized. If not, the influence of each modality is strongly dependent on its scores.

CombSUM with scores was used in Jones et al (2004); Chevallet et al (2005); Martín-Valdivia et al (2005) and was used only once based on rank in El Demerdash et al (2007). Similarly to Mulhem (2008); Broda et al (2009), combSUM is applied if and only if the visual score is above a given threshold based on TF–IDF value for images annotations in Navarro et al (2008c,b,d,a, 2009).

Borda Count

The Borda count election method was developed in the political context in 1770 to create a ranked list of candidates. Each voter ranks all candidates and the sum of the ranks for all voters determines the score of each candidate from which a final ranking can be derived. This method was applied in information fusion in Ho et al (1994); van Erp and Schomaker (2000) and in ImageCLEF in Overell et al (2008). Borda count is strictly equivalent to combSUM on ranks.

6.3.2.5 CombMNZ

A variant of the combSUM method is the combMNZ combination rule which aims at giving more importance to the documents retrieved by several systems as follows (Shaw and Fox, 1994):

$$S_{mixed}(d) = F(d) \sum_{j=1}^{N_j} S_j(d), \tag{6.14}$$

where $F(d)$ is equal to the number of systems that retrieved d. CombMNZ was slightly modified by Inkpen et al (2008) for the photo retrieval task where a weight was applied to the normalized scores of each modality in order to control their respective influences.

6.3.2.6 CombMAX and CombMIN

Contrary to combSUM, the combMAX and combMIN rules put all their confidence in one single modality as follows:

$$\text{combMAX:} \qquad S_{mixed}(d) = \arg \max_{j=1:N_j} (S_j(d)), \tag{6.15}$$

$$\text{combMIN:} \qquad S_{mixed}(d) = \arg \min_{j=1:N_j} (S_j(d)). \tag{6.16}$$

CombMAX and combMIN were used both for photo and medical image retrieval by Besançon and Millet (2005); Chevallet et al (2005); Villena-Román et al (2007b,a) using normalized scores. CombMIN based on ranks was used in Ferecatu and Sahbi (2008) and is similar to combMAX based on score.

A hybrid rule based both combMAX and combMIN is proposed by Villena-Román et al (2007b,a):

$$S_{mixed}(d) = \text{combMAX}(S_j(d)) + \frac{\text{combMIN}^2(S_j(d))}{\text{combMAX}(S_j(d)) + \text{combMIN}(S_j(d))}. \tag{6.17}$$

It allows importance to be given to the minimum scores only if the latter has sufficiently high values.

6.3.2.7 CombPROD

The combPROD combination rule uses the product of scores to compute S_{mixed}:

$$S_{mixed}(d) = \prod_{j=1}^{N_j} (S_j(d)). \tag{6.18}$$

CombPROD favors documents with high scores in all modalities and was used for both photo and medical image retrieval by Martínez-Fernández et al (2004).

6.3.3 Inter–media Feedback with Query Expansion

The idea of query expansion is to modify the original query based on either available documents in the database or given rules (i.e. use of synonyms of query terms) with an aim of guessing the user's intentions. It was successfully applied to TREC[1] test collections in Belkin et al (1993), and Saracevic and Kantor (1988) states explicitly that taking into account the different results of the formulations could lead to retrieval performance better than that of any of the individual query formulations.

Query expansion was widely used in ImageCLEF and particularly for fusing textual and visual information where one modality provides a feedback to the other by means of query expansion, which is commonly called inter–media feedback in ImageCLEF (El Demerdash et al, 2009b).

6.3.3.1 Textual Query Expansion

Inter–media feedback query expansion is based on textual query expansion in most of the papers. Typically textual annotations from the top visually–ranked images (or from a mixed run) are used to expand a textual query (Ruiz and Srikanth, 2004; Müller et al, 2004; Besançon et al, 2004; Jones and McDonald, 2005; Chang et al, 2005; Maillot et al, 2006; Jair Escalante et al, 2007; Chang and Chen, 2007; Torjmen et al, 2007; Gao et al, 2007; Yamauchi et al, 2008; Gao et al, 2008; El Demerdash et al, 2008; Navarro et al, 2008c,b; Chang and Chen, 2008; El Demerdash et al, 2009a; Navarro et al, 2009).

Alternatively, text–based queries are built based on the automatically detected concepts present in the query image in Jair Escalante et al (2007); Tollari et al (2008); Inoue and Grover (2008); Popescu et al (2008).

In Kalpathy-Cramer et al (2008), the medical image modality (x–ray, computed tomography, etc.) is automatically detected from visual features and used as query expansion for text–based retrieval.

6.3.3.2 Visual Query Expansion

A less common approach for inter–media query expansion is proposed by Benczúr et al (2007), where the regions of images that are correlated with the title of the topic are used as visual queries with a CBIR engine.

[1] Text REtrieval Conference (TREC, http://trec.nist.gov/)

6.3.4 Other Approaches

Some of the techniques used in ImageCLEF for fusing textual and visual information do not correspond to any of the above–mentioned categories and proposed innovative approaches for merging information sources.

A simple approach is proposed by Radhouani et al (2009) who use visual features to detect the imaging modality in a first step. Then, images returned by a TBIR engine are filtered according to the modality of the query image.

A word–image ontology based on images retrieved by Google images using all nouns contained in the WordNet ontology is used by Chang and Chen (2006); Lacoste et al (2006). The textual query is mapped to a visual query based on the word–image ontology, which is then submitted to a CBIR system to obtain a final list of images.

Two innovative reordering methods based on ranks and applied to subgroups of documents are proposed by Myoupo et al (2009). In the first approach, the combSUM rule is iteratively applied on groups of documents within the lists, where groups are created using a sliding window consisting of groups *N* consecutive documents within each list. The second merging strategy is based on homogeneous blocks as follows: in the list of text retrieved documents, images are clustered according to their visual similarities to create blocks. Then, blocks are reordered among them according to their internal mean scores.

6.3.5 Overview of the Methods from 2004–2009

An overview of the main techniques and their interdependences is proposed in Figure 6.1. The late fusion techniques are most widely used and developed. The distribution of the various fusion approaches is detailed in Figure 6.2. It is important to note that some groups used a combination of the fusion techniques (see Maillot et al (2006)) and often research groups reused their techniques with slight modifications from one year to another and across tasks, which potentially exaggerates the trends in Figure 6.2.

6.4 Justification for the Approaches and Generally Known Problems

In this section, the justification of the methods, identified trends as well as lessons learned from seven years of multi–modal image retrieval are discussed.

Figures 6.1 and 6.2 clearly show three different choices of the system level for combining the modalities: at the input level with query expansion, internally with early fusion and at the output level with late fusion. Merging modalities at the input

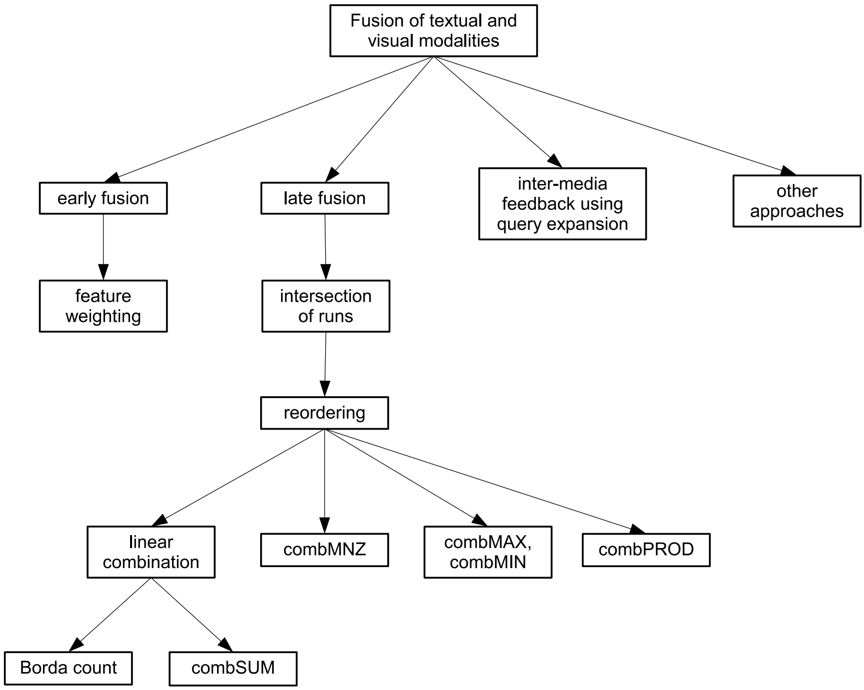

Fig. 6.1: Overview of the techniques.

level with query expansion techniques aims at improving the recall as the additional keywords (or query images) enable it to retrieve more potentially relevant images, but also involve the risk of proposing too many results to the user and thereby decreasing the precision. Early fusion enables a comprehensive overview of the multi–modal information by combining modalities inside the IR system and offers potentially high flexibility for promoting relevant modalities in the context of a particular query. Unfortunately, it is difficult to put into practice because it relies on large and heterogeneous feature spaces that become less distinctive, due to what is called the curse of dimensionality. Moreover, combining binary and categorical variable that are textual attributes with continuous and correlated visual features is not trivial and negative interactions among features can occur (see (Bell, 2003)). Consequently, it was shown to perform very well when textual features are combined with a small number of basic visual features such as in Berber and Alpkoçak (2009), which obtained best performance in last year's (2009) medical image retrieval task. Late fusion techniques are by far the most frequently utilized with more than 60% of the papers dealing with textual and information fusion. This is not surprising as late fusion allows for a straightforward combination of any system

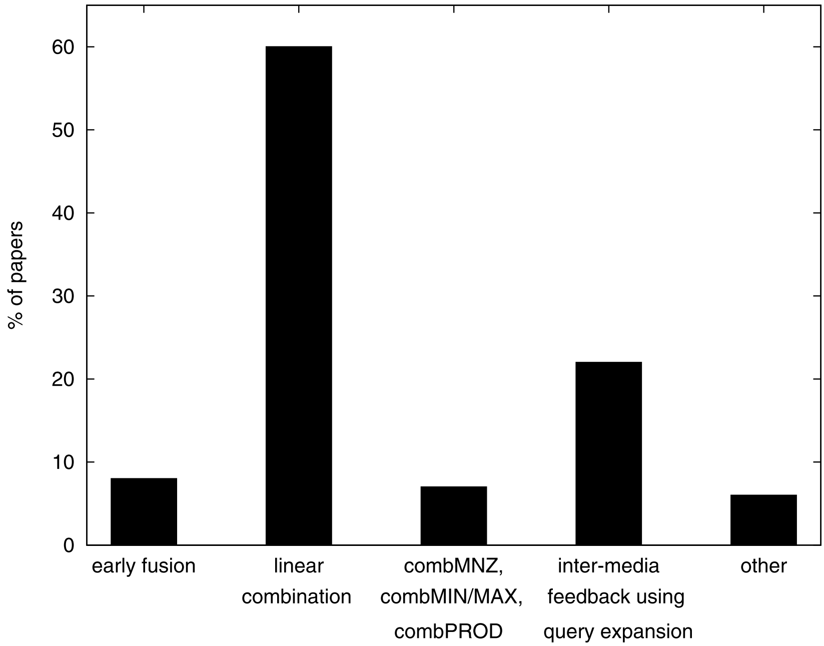

Fig. 6.2: Distribution of fusion approaches.

delivering a ranked list of documents. Most of the research groups focused on the performance of each independent system, which is a necessary condition to achieve high mixed performance.

When both TBIR and CBIR achieve acceptable performance, the choice of the fusion technique should rely on the analysis of the trends of each independent system as well as their complementarity and relevance to the image retrieval task (see (Zhou et al, 2010)). For instance, the combMAX combination rule favors the documents that are highly ranked in one system ('Dark Horse effect', (Vogt and Cottrell, 1999)) and is thus not robust to errors. On the other hand, combSUM and combMNZ favor the documents widely returned to minimize the errors ('Chorus effect') but relevant documents can obtain high ranks even if they are returned by few systems. Nevertheless, some of the approaches have fundamental limitations. This is the case with the linear combination using fixed weight for each document, as it puts blind confidence in one of the modalities and banishes the other one. This is not desirable as each modality usually behaves differently with each query and each set of documents. Consequently, late fusion techniques able to foster the modality with higher confidence are preferable as they allow the selection of the appropriate modality based on the query and the database. The idea of fostering the modality with confidence was found in various approaches such as combPROD or when linear combinations of scores are applied only if the scores of each modality are above a given threshold. Interestingly, Myoupo et al (2009) showed that the reordering of documents was much more adapted when carried out within subgroups of document instead of global reordering.

Several studies tried to enhance the diversity of the retrieved documents using mixed retrieval (see (Chang and Chen, 2008; Ah-Pine et al, 2008; Hare et al, 2009; Zuccon et al, 2009)), which was often based on cross–modality clustering (see (Arni et al, 2008; Lestari Paramita et al, 2009)). This was promoted by the organizers starting from 2008 for the photo retrieval task.

Finally, a quantitative comparison of the various fusion techniques was difficult to perform as the retrieval performance strongly depends on the performance of each independent IR system, which varied significantly among research groups. It was observed that mixed runs achieve better performance than single modalities in most of the cases. Most often, a degradation of performance is observed when the CBIR system achieves poor performances such as in Boutsis and Kalamboukis (2009).

6.5 Conclusions

In this chapter, the various approaches used during the past seven years in the ImageCLEF campaign were reviewed. Clear trends among techniques have been identified and discussed. A major observation is that CBIR systems have become mature enough to extract semantic information that is complementary to textual information, thus allowing enhancement of the quality of retrieval both in terms of precision and recall. However it was observed that combining textual and visual information is not devoid of risks and can degrade the retrieval performance if the fusion technique is not adapted to the information retrieval paradigm as well as to the TBIR and CBIR systems used. The key to using data fusion techniques is making the most of both textual and visual modalities.

Acknowledgements This work was supported by the Swiss National Science Foundation (FNS) with grant 200020–118638/1 and 205321–127355/9.

References

Adriani M, Framadhan A (2005) The University of Indonesia's participation in ImageCLEF 2005. In: Working Notes of CLEF 2005, Vienna, Austria

Ah-Pine J, Cifarelli C, Clinchant S, Renders JM (2008) XRCE's participation to ImageCLEF 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Ah-Pine J, Clinchant S, Csurka G, Liu Y (2009) XRCE's participation in ImageCLEF 2009. In: Working Notes of CLEF 20098, Corfu, Greece

Alvarez C, Id Oumohmed A, Mignotte M, Nie JY (2004) Toward cross–language and cross–media image retrieval. In: Working Notes of CLEF 2004, Bath, England

Arni T, Clough PD, Sanderson M, Grubinger M (2008) Overview of the ImageCLEFphoto 2008 photographic retrieval task. In: Working Notes of CLEF 2008, pp 500–511
Belkin NJ, Cool C, Croft WB, Callan JP (1993) The effect of multiple query representations on information retrieval system performance. In: Korfhage R, Rasmussen EM, Willett P (eds) Proceedings of the 16st annual international ACM SIGIR conference on research and development in information retrieval. ACM press, Pittsburgh, PA, USA, pp 339–346
Belkin NJ, Kantor P, Cool C, Quatrain R (1994) Combining evidence for information retrieval. In: TREC–2: The Second Text REtrieval Conference, pp 35–44
Bell AJ (2003) The co–information lattice. In: Proceedings of the 4th international symposium on independent component analysis and blind signal separation (ICA2003). Springer, Nara, Japan, pp 921–926
Benczúr A, Bíró I, Brendel M, Csalogány K, Daróczy B, Siklósi D (2007) Cross–modal retrieval by text and image feature biclustering. In: Working Notes of CLEF 2007, Budapest, Hungary
Berber T, Alpkoçak A (2009) DEU at ImageCLEFmed 2009: Evaluating re–ranking and integrated retrieval model. In: Working Notes of CLEF 2009, Corfu, Greece
Besançon R, Millet C (2005) Merging results from different media: Lic2m experiments at ImageCLEF 2005. In: Working Notes of CLEF 2005, Vienna, Austria
Besançon R, Hède P, Moëllic PA, Fluhr C (2004) LIC2M experiments at ImageCLEF 2004. In: Working Notes of CLEF 2004, Bath, England
Boutsis I, Kalamboukis T (2009) Combined content–based and semantic image retrieval. In: Working Notes of CLEF 2009, Corfu, Greece
Broda B, Paradowski M, Kwanicka H (2009) Multimodal photo retrieval through finding similar documents enhanced with visual clues — a baseline method. In: Working Notes of CLEF 2009, Corfu, Greece
Chang YC, Chen HH (2006) Approaches of using a word–image ontology and an annotated image corpus as intermedia for cross–language image retrieval. In: Working Notes of CLEF 2006, Alicante, Spain
Chang YC, Chen HH (2007) Experiment for using web information to do query and document expansion. In: Working Notes of the CLEF 2007, Budapest, Hungary
Chang YC, Chen HH (2008) Increasing relevance and diversity in photo retrieval by result fusion. In: Working Notes of CLEF 2008, Aarhus, Denmark
Chang YC, Lin WC, Chen HH (2005) Combing text and image queries at ImageCLEF2005. In: Working Notes of CLEF 2005, Vienna, Austria
Cheng PC, Chien BC, Ke HR, Yang WP (2004) KIDS's evaluation in the medical image retrieval task at ImageCLEF 2004. In: Working Notes of CLEF 2004, Bath, England
Cheng PC, Yeh JY, Ke HR, Chien BC, Yang WP (2004) NCTU–ISU's evaluation for the user–centered search task at ImageCLEF 2004. In: Working Notes of CLEF 2004, Bath, England
Cheng PC, Chien BC, Ke HR, Yang WP (2005) NCTU_DBLAB@ImageCLEFmed 2005: Medical image retrieval task. In: Working Notes of CLEF 2005, Vienna, Austria
Chevallet JP, Lim JH, Radhouani S (2005) Using ontology dimensions and negative expansion to solve precise queries in the CLEF medical task. In: Working Notes of the CLEF 2005, Vienna, Austria
Clinchant S, Renders JM, Csurka G (2007) XRCE's participation to ImageCLEFphoto 2007. In: Working Notes of CLEF 2007, Budapest, Hungary
Daróczy B, Petrás I, Benczúr AA, Zsolt Fekete Z, Nemeskey D, Siklósi D, Weiner Z (2009) SZTAKI @ ImageCLEF 2009. In: Working Notes of CLEF 2009, Corfu, Greece
Daumke P, Paetzold J, Markó K (2006) Morphosaurus in ImageCLEF 2006: The effect of subwords on biomedical IR. In: Working Notes of CLEF 2006, Alicante, Spain
Depeursinge A, Racoceanu D, Iavindrasana J, Cohen G, Platon A, Poletti PA, Müller H (2010–to appear) Fusing visual and clinical information for lung tissue classification in HRCT data. Journal of Artificial Intelligence in Medicine
Deselaers T, Weyand T, Keysers D, Macherey W, Ney H (2005) FIRE in ImageCLEF 2005: Combining content–based image retrieval with textual information retrieval. In: Working Notes of the CLEF Workshop, Vienna, Austria

Deselaers T, Weyand T, Ney H (2006) Image retrieval and annotation using maximum entropy. In: Working Notes of CLEF 2006, Alicante, Spain

Deselaers T, Gass T, Weyand T, Ney H (2007) FIRE in ImageCLEF 2007. In: Working Notes of CLEF 2007, Budapest, Hungary

Díaz-Galiano MC, García-Cumbreras MA, Martín-Valdivia MT, Montejo-Raez A, Ureña-López LA (2006) SINAI at ImageCLEF 2006. In: Working Notes of CLEF 2006, Alicante, Spain

Díaz-Galiano MC, García-Cumbreras MA, Martín-Valdivia MT, Montejo-Raez A, Ureña-López LA (2007) SINAI at ImageCLEF 2007. In: Working Notes of CLEF2007, Budapest, Hungary

Díaz-Galiano MC, García-Cumbreras MA, Martín-Valdivia MT, Ureña-López LA, Montejo-Raez A (2008) SINAI at ImageCLEFmed 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Douze M, Guillaumin M, Mensink T, Schmid C, Verbeek J (2009) INRIA–LEARs participation to ImageCLEF 2009. In: Working Notes of CLEF 2009, Corfu, Greece

El Demerdash O, Kosseim L, Bergler S (2007) Experiments with clustering the collection at ImageCLEF 2007. In: Working Notes of CLEF 2007, Budapest, Hungary

El Demerdash O, Kosseim L, Bergler S (2008) CLaC at ImageCLEFPhoto 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

El Demerdash O, Bergler S, Kosseim L (2009a) CLaC at ImageCLEF 2009. In: Working Notes of CLEF 2009, Corfu, Greece

El Demerdash O, Kosseim L, Bergler S (2009b) Image retrieval by inter–media fusion and pseudo–relevance feedback. In: Evaluating systems for multilingual and multimodal information access, pp 605–611

van Erp M, Schomaker L (2000) Variants of the borda count method for combining ranked classifier hypotheses. In: Seventh International Workshop on Frontiers in Handwriting Recognition, pp 443–452

Fakeri-Tabrizi A, Amini MR, Tollari S, Gallinari P (2008) UPMC/LIP6 at ImageCLEF's WikipediaMM: An image–annotation model for an image search–engine. In: Working Notes of CLEF 2008, Aarhus, Denmark

Ferecatu M, Sahbi H (2008) TELECOM ParisTech at ImageClefphoto 2008: Bi–modal text and image retrieval with diversity enhancement. In: Working Notes of CLEF 2008, Aarhus, Denmark

Florea F, Rogozan A, Cornea V, Bensrhair A, Darmoni S (2006) MedIC/CISMeF at ImageCLEF 2006: Image annotation and retrieval tasks. In: Working Notes of CLEF 2006, Alicante, Spain

Frank Hsu D, Taksa I (2005) Comparing rank and score combination methods for data fusion in information retrieval. Information Retrieval 8(3):449–480

Gao S, Lim JH (2009) I2R at ImageCLEF photo retrieval 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Gao S, Chevallet JP, Le THD, Pham TT, Lim JH (2007) IPAL at ImageClef 2007 mixing features, models and knowledge. In: Working Notes of CLEF 2007, Budapest, Hungary

Gao S, Chevallet JP, Lim JH (2008) IPAL at CLEF 2008: Mixed–modality based image search, novelty based re–ranking and extended matching. In: Working Notes of CLEF 2008, Aarhus, Denmark

Glotin H, Zhao Z (2008) Affnity propagation promoting diversity in visuo–entropic and text features for CLEF Photo retrieval 2008 campaign. In: Working Notes of CLEF 2008, Aarhus, Denmark

Gobeill J, Müller H, Ruch P (2006) Query and document translation by automatic text categorization: A simple approach to establish a strong textual baseline for ImageCLEFmed 2006. In: Working Notes of CLEF 2006, Alicante, Spain

Granados R, Benavent X, García-Serrano A, Goñi JM (2008) MIRACLE–FI at ImageCLEFphoto 2008: Experiences in merging text–based and content–based retrievals. In: Working Notes of CLEF 2008, Aarhus, Denmark

Granados R, Benavent X, Agerri R, García-Serrano A, Goñi JM, Gomar J, de Ves E, Domingo J, Ayala G (2009) MIRACLE (FI) at ImageCLEFphoto 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Gunes H, Piccardi M (2005) Affect recognition from face and body: early fusion vs. late fusion. In: 2005 IEEE International Conference on Systems, Man and Cybernetics, vol 4. IEEE Computer Society, Big Island, Hawaii, pp 3437–3443

Hare JS, Dupplaw DP, Lewis PH (2009) IAM@ImageCLEFphoto 2009: Experiments on maximising diversity using image features. In: Working Notes of CLEF 2009, Corfu, Greece

Hersh W, Kalpathy-Cramer J, Jensen J (2006) Medical image retrieval and automated annotation: OHSU at ImageCLEF 2006. In: Working Notes of CLEF 2006, Alicante, Spain

Hersh W, Kalpathy-Cramer J, Jensen J (2007) Medical image retrieval and automated annotation: OHSU at ImageCLEF 2006. In: Peters C, D. PC, Gey FC, Karlgren J, Magnini B, Oard DW, de Rijke M, Stempfhuber M (eds) CLEF. Lecture Notes in Computer Science (LNCS), vol 4730. Springer, pp 660–669

Ho TK, Hull JJ, Srihari SN (1994) Decision combination in multiple classifier systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 16:66–75

Hoi SCH (2007) Cross–language and cross–media image retrieval: An empirical study at ImageCLEF2007. In: Working Notes of CLEF 2007, Budapest, Hungary

Hoi SCH, Zhu J, Lyu MR (2005) CUHK experiments with ImageCLEF 2005. In: Working Notes of CLEF 2005, Vienna, Austria

Inkpen D, Stogaitis M, DeGuire F, Alzghool M (2008) Clustering for photo retrieval at ImageCLEF 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Inoue M, Grover P (2008) Effects of visual concept–based post–retrieval clustering in ImageCLEFphoto 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Jair Escalante H, Hernández CA, López A, Marín HM, Montes M, Morales E, Sucar LE, Villaseñor L (2007) TIA–INAOE's participation at ImageCLEF 2007. In: Working Notes of CLEF 2007, Budapest, Hungary

Jair Escalante H, Gonzáles JA, Hernández CA, López A, Montes M, Morales E, Sucar LE, Villaseñor L (2008) TIA–INAOE's participation at ImageCLEF 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Jair Escalante H, Gonzáles JA, Hernández CA, López A, Montes M, Morales E, Ruiz E, Sucar LE, Villaseñor L (2009) TIA–INAOE's participation at ImageCLEF 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Järvelin A, Wilkins P, Adamek T, Airio E, Jones GJF, Smeaton AF, Sormunen E (2007) DCU and UTA at ImageCLEFphoto 2007. In: Working Notes of CLEF 2007, Budapest, Hungary

Jensen JR, Hersh WR (2005) Manual query modification and automatic translation to improve cross–language medical image retrieval. In: Working Notes of CLEF 2005, Vienna, Austria

Jones GJF, McDonald K (2005) Dublin City University at CLEF 2005: Experiments with the ImageCLEF St Andrew's collection. In: Working Notes of CLEF 2005, Vienna, Austria

Jones GJF, Groves D, Khasin A, Lam-Adesina A, Mellebeek B, Way A (2004) Dublin City University at CLEF 2004: Experiments with the ImageCLEF St Andrew's collection. In: Working Notes of CLEF 2004, Bath, England

Kalpathy-Cramer J, Hersh W (2007) Medical image retrieval and automatic annotation: OHSU at ImageCLEF 2007. In: Working Notes of CLEF 2007, Budapest, Hungary

Kalpathy-Cramer J, Bedrick S, Hatt W, Hersh W (2008) Multimodal medical image retrieval: OHSU at ImageCLEF 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Lacoste C, Chevallet JP, Lim JH, Xiong W, Racoceanu D (2006) IPAL knowledge–based medical image retrieval in ImageCLEFmed 2006. In: Working Notes of CLEF 2006, Alicante, Spain

Lee JH (1995) Combining multiple evidence from different properties of weighting schemes. In: Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval. ACM press, pp 180–188

Lee JH (1997) Analyses of multiple evidence combination. In: Proceedings of the 20th annual international ACM SIGIR conference on research and development in information retrieval. ACM press, New York, NY, USA, pp 267–276

Lestari Paramita M, Sanderson M, Clough PD (2009) Diversity in photo retrieval: Overview of the ImageCLEFPhoto task 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Lim JH, Chevallet JP (2005) A structured learning approach for medical image indexing and retrieval. In: Working Notes of CLEF 2005, Vienna, Austria

Lin WC, Chang YC, Chen HH (2004) From text to image: Generating visual query for image retrieval. In: Working Notes of CLEF 2004, Bath, England

Magalhães J, Overell S, Rüger S (2007) Exploring image, text and geographic evidences in ImageCLEF 2007. In: Working Notes of CLEF 2007, Budapest, Hungary

Maillot N, Chevallet JP, Valea V, Lim JH (2006) IPAL inter–media pseudo–relevance feedback approach to ImageCLEF 2006 photo retrieval. In: Working Notes of CLEF 2006, Alicante, Spain

Martín-Valdivia MT, García-Cumbreras MA, Díaz-Galiano MC, Ureña-López LA, Montejo-Raez A (2005) SINAI at ImageCLEF 2005. In: Working Notes of CLEF 2005, Vienna, Austria

Martínez-Fernández JL, Serrano AG, Villena-Román J, Sáenz VDM, Tortosa SG, Castagnone M, Alonso J (2004) MIRACLE at ImageCLEF 2004. In: Working Notes of CLEF 2004, Bath, England

Montague M, Aslam JA (2001) Relevance score normalization for metasearch. In: Proceedings of the tenth international conference on information and knowledge management. ACM press, pp 427–433

Moulin C, Barat C, Géry M, Ducottet C, Largeron C (2008) UJM at ImageCLEFwiki 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Moulin C, Barat C, Lemaître C, Géry M, Ducottet C, Largeron C (2009) Combining text/image in WikipediaMM task 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Mulhem P (2008) LIG at ImageCLEFphoto 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Mulhem P, Chevallet JP, Quénot G, Al Batal R (2009) MRIM–LIG at ImageCLEF 2009: Photo retrieval and photo annotation tasks. In: Working Notes of CLEF 2009, Corfu, Greece

Müller H, Geissbuhler A, Ruch P (2004) Report on the CLEF experiment: Combining image and multi–lingual search for medical image retrieval. In: Working Notes of CLEF 2004, Bath, England

Müller H, Geissbuhler A, Marty J, Lovis C, Ruch P (2005) Using medGIFT and easyIR for the ImageCLEF 2005 evaluation tasks. In: Working Notes of CLEF 2005, Vienna, Austria

Müller H, Deselaers T, Kim E, Kalpathy-Cramer J, Deserno TM, Clough PD, Hersh W (2008) Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks. In: CLEF 2007 Proceedings. Lecture Notes in Computer Science (LNCS), vol 5152. Springer, Budapest, Hungary, pp 473–491

Myoupo D, Popescu A, Le Borgne H, Moëllic PA (2009) Visual reranking for image retrieval over the Wikipedia corpus. In: Working Notes of CLEF 2009, Corfu, Greece

Navarro S, Díaz MC, Muñoz R, García MA, Llopis F, Martín MT, Ureña-López LA, Montejo-Raez A (2008a) Text–mess in the medical retrieval ImageCLEF08 task. In: Working Notes of CLEF 2008, Aarhus, Denmark

Navarro S, García MA, Llopis F, Díaz MC, Muñoz R, Martín MT, Ureña-López LA, Montejo-Raez A (2008b) Text–mess in the ImageCLEFphoto task. In: Working Notes of CLEF 2008, Aarhus, Denmark

Navarro S, Llopis F, Muñoz R (2008c) Different multimodal approaches using IR–n in ImageCLEFphoto 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Navarro S, Muñoz R, Llopis F (2008d) A multimodal approach to the medical retrieval task using IR–n. In: Working Notes of CLEF 2008, Aarhus, Denmark

Navarro S, Muñoz R, Llopis F (2009) Evaluating fusion techniques at different domains at ImageCLEF subtasks. In: Working Notes of CLEF 2009, Corfu, Greece

O'Hare N, Wilkins P, Gurrin C, Newman E, Jones GJF, Smeaton AF (2008) DCU at ImageCLEFPhoto 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Overell S, Llorente A, Liu H, Hu R, Rae A, Zhu J, Song D, Rüger S (2008) MMIS at ImageCLEF 2008: Experiments combining different evidence sources. In: Working Notes of CLEF 2008, Aarhus, Denmark

Popescu A, Le Borgne H, Moëllic PA (2008) Conceptual image retrieval over the Wikipedia corpus. In: Working Notes of CLEF 2008, Aarhus, Denmark

Quinlan RJ (1986) Induction of decision trees. Machine Learning 1(1):81–106

Rácz S, Daróczy B, Siklósi D, Pereszlényi A, Brendel M, Benczúr A (2008) Increasing cluster recall of cross–modal image retrieval. In: Working Notes of CLEF 2008, Aarhus, Denmark

Radhouani S, Kalpathy-Cramer J, Bedrick S, Bakke B, Hersh W (2009) Multimodal medical image retrieval improving precision at ImageCLEF 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Rahman MM, Sood V, Desai BC, Bhattacharya P (2006) CINDI at ImageCLEF 2006: Image retrieval & annotation tasks for the general photographic and medical image collections. In: Working Notes of CLEF 2006, Alicante, Spain

Rahman MM, Desai BC, Bhattacharya P (2007) Multi–modal interactive approach to ImageCLEF 2007 photographic and medical retrieval tasks by CINDI. In: Working Notes of CLEF 2007, Budapest, Hungary

Ruiz ME (2009) UNT at ImageCLEFmed 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Ruiz ME, Southwick SB (2005) UB at CLEF 2005: Medical image retrieval task. In: Working Notes of CLEF 2005, Vienna, Austria

Ruiz ME, Srikanth M (2004) UB at CLEF2004: Part 2 — cross language medical image retrieval. In: Working Notes of CLEF 2004, Bath, England

Salton G, Buckley C (1988) Term weighting approaches in automatic text retrieval. Information Processing and Management 24(5):513–523

Saracevic T, Kantor P (1988) A study of information seeking and retrieving. ii. users, questions, and effectiveness. Journal of the American Society for Information Science 39:177–196

Shaw JA, Fox EA (1994) Combination of multiple searches. In: TREC–2: The Second Text REtrieval Conference, pp 243–252

Simpson M, Rahman MM, Demner-Fushman D, Antani S, Thoma GR (2009) Text– and content–based approaches to image retrieval for the ImageCLEF 2009 medical retrieval track. In: Working Notes of CLEF 2009, Corfu, Greece

Smeulders AWM, Worring M, Santini S, Gupta A, Jain R (2000) Content–based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12):1349–1380

Snoek CGM, Worring M, Smeulders AWM (2005) Early versus late fusion in semantic video analysis. In: Proceedings of the 13th annual ACM international conference on Multimedia. ACM press, New York, NY, USA, pp 399–402

Tollari S, Detyniecki M, Fakeri-Tabrizi MR, Ali ad Amini, Gallinari P (2008) UPMC/LIP6 at ImageCLEFphoto 2008: on the exploitation of visual concepts (VCDT). In: Working Notes of CLEF 2008, Aarhus, Denmark

Torjmen M, Pinel-Sauvagnat K, Boughanem M (2007) Using pseudo–relevance feedback to improve image retrieval results. In: Working Notes of CLEF 2007, Budapest, Hungary

Torjmen M, Pinel-Sauvagnat K, Boughanem M (2008) Methods for combining content–based and textual–based approaches in medical image retrieval. In: Working Notes of CLEF 2008, Aarhus, Denmark

Villena-Román J, González-Cristóbal JC, Goñi-Menoyo JM, Martínez-Fernández JL, Fernández JJ (2005) MIRACLE's combination of visual and textual queries for medical images retrieval. In: Working Notes of CLEF 2005, Vienna, Austria

Villena-Román J, Lana-Serrano S, González-Cristóbal JC (2007a) MIRACLE at ImageCLEFmed 2007: Merging textual and visual strategies to improve medical image retrieval. In: Working Notes of CLEF 2007, Budapest, Hungary

Villena-Román J, Lana-Serrano S, Martínez-Fernández JL, González-Cristóbal JC (2007b) MIRACLE at ImageCLEFphoto 2007: Evaluation of merging strategies for multilingual and multimedia information retrieval. In: Working Notes of CLEF 2007, Budapest, Hungary

Vogt CC, Cottrell GW (1999) Fusion via a linear combination of scores. Information Retrieval 1(3):151–173

Wilhelm T, Eibl M (2006) ImageCLEF 2006 experiments at the Chemnitz Technical University. In: Working Notes of CLEF 2006, Alicante, Spain

Wilhelm T, Kürsten J, Eibl M (2007) Experiments for the ImageCLEF 2007 photographic retrieval task. In: Working Notes of CLEF 2007, Budapest, Hungary

Yamauchi K, Nomura T, Usui K, Kamoi Y, Eto M, Takagi T (2008) Meiji University at ImageCLEF2008 photo retrieval task: Evaluation of image retrieval methods integrating different media. In: Working Notes of CLEF 2008, Aarhus, Denmark

Ye Z, Huang X, Lin H (2009) Towards a better performance for medical image retrieval using an integrated approach. In: Working Notes of CLEF 2009, Corfu, Greece

van Zaanen M, de Croon G (2004) FINT: Find Images aNd Text. In: Working Notes of CLEF 2004, Bath, England

Zhao Z, Glotin H (2008) Concept content based Wikipedia WEB image retrieval using CLEF VCDT 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Zhou X, Gobeill J, Ruch P, Müller H (2007) University and Hospitals of Geneva at ImageCLEF 2007. In: Working Notes of CLEF 2007, Budapest, Hungary

Zhou X, Gobeill J, Müller H (2008a) MedGIFT at ImageCLEF 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Zhou X, Eggel I, Müller H (2009) The MedGIFT group at ImageCLEF 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Zhou X, Depeursinge A, Müller H (2010) Information fusion for combining visual and textual image retrieval. In: Pattern Recognition, International Conference on. IEEE Computer Society, Istambul, Turkey

Zhou Z, Tian Y, Li Y, Liu T, Huang T, Gao W (2008b) PKU at ImageCLEF 2008: Experiments with query extension techniques for text–based and content–based image retrieval. In: Working Notes of CLEF 2008, Aarhus, Denmark

Zuccon G, Leelanupab T, Goyal A, Halvey M, Punitha P, Jose JM (2009) The University of Glasgow at ImageClefPhoto 2009. In: Working Notes of CLEF 2009, Corfu, Greece