

# The Scholarly Impact of TRECVID (2003-2009)

(pre-print)

## Clare V. Thornley

School of Information and Library Studies, University College Dublin, Belfield, Dublin 4, Ireland. E-Mail: [clare.thornley@ucd.ie](mailto:clare.thornley@ucd.ie)

## Andrea C. Johnson

School of Information and Library Studies, University College Dublin, Belfield, Dublin 4, Ireland. E-Mail: [andrea.johnson@ucd.ie](mailto:andrea.johnson@ucd.ie)

## Alan F. Smeaton

CLARITY: Centre for Sensor Web Technologies, School of Computing, Dublin City University, Glasnevin, Dublin 9, Ireland. E-Mail: [alan.smeaton@dcu.ie](mailto:alan.smeaton@dcu.ie)

## Hyowon Lee

CLARITY: Centre for Sensor Web Technologies, School of Computing, Dublin City University, Glasnevin, Dublin 9, Ireland. E-Mail: [hlee@computing.dcu.ie](mailto:hlee@computing.dcu.ie)

## Abstract

This paper reports on an investigation into the scholarly impact of the TRECVID (TREC Video Retrieval Evaluation) benchmarking conferences between 2003 and 2009. The contribution of TRECVID to research in video retrieval is assessed by analyzing publication content to show the development of techniques and approaches over time and by analyzing publication impact through publication numbers and citation analysis. Popular conference and journal venues for TRECVID publications are identified in terms of number of citations received. For a selection of participants at different career stages, the relative importance of TRECVID publications in terms of citations *vis a vis* their other publications is investigated. TRECVID, as an evaluation conference, provides data on which research teams ‘scored’ highly against the evaluation criteria and the relationship between ‘top scoring’ teams at TRECVID and the ‘top scoring’ papers in terms of citations is analysed. A strong relationship was found between ‘success’ at TRECVID and ‘success’ at citations both for high scoring and low scoring teams. The implications of the study in terms of the value of TRECVID as a research activity, and the value of bibliometric analysis as a research evaluation tool, are discussed.

## Introduction

In this paper we report on results from a study to investigate the scholarly impact of the annual series of TRECVID benchmarking conferences. TRECVID started out as one of several tracks in the larger TREC (Text Retrieval and Evaluation Conference) benchmarking conference series in 2001, and it became a separate activity in 2003. The overall aim of TREC and TRECVID is to provide access to large scale test collections so

that newly developed techniques for content-based operations like search can be tested and compared in an open, metrics-based way and in this way help to progress the field of information retrieval (IR). TRECVideo uses the same model of evaluation as TREC but the focus is on techniques for digital video whereas TREC focuses on text derived from documents, web pages, blogs, automatic speech recognition, etc. After almost 20 years of activity, the TREC model of evaluation is not without its critics and discussion (Robertson, 2008) on its reliability and, in particular, its validity are widespread in the literature. It does, however, provide the only forum for the large scale testing of new IR techniques.

Our study assesses the scholarly impact of TRECVideo and by this we mean the extent to which the TRECVideo conferences have influenced the development of new thinking and techniques in the field of video retrieval. After 7 years as a standalone benchmarking conference and 2 years as a TREC track it is reasonable to ask whether it has been a successful forum for the development of improved techniques. In a broader sense has TRECVideo been successful in developing new ideas and approaches to the problems in video management? In this study investigate these questions using bibliometric tools examining the scientific publications written as result of TRECVideo and their associated citations. This builds upon existing work done by NIST in 2010 on the economic impact study of TREC (Rowe *et al.*, 2010) investigating the economic success of the main TREC activity in developing new IR techniques. It reached some, broadly positive, conclusions regarding the financial return on investment of funding TREC over a 19 year period but it did not examine the scholarly or academic impact of the conference.

What exactly is scholarly impact? How can we measure the effect or impact that TRECVideo has had on video retrieval research both in terms of thinking and practice? How far has the influence of this annual benchmarking conference spread? We can, to an extent, answer some of these questions by examining the number of publications derived from the benchmarking activity in TRECVideo and the number of citations they have received. We defined these as publications that could not have been written if TRECVideo hadn't happened because of their reliance in some way on TRECVideo data and/or the TRECVideo benchmarking process. TRECVideo and TREC in general are different from the majority of conferences because they enable research in the specific way of providing an evaluation and benchmarking process as well as providing a forum for disseminating research. The TRECVideo conference itself is just the final stage of a year-long evaluation process which only participants have access to. The research on new video retrieval techniques could have been done without involvement in the TRECVideo benchmarking. It would be very difficult, however, for researchers to evaluate and compare their results against other possible approaches.

We examine the papers derived from TRECVideo and investigate where they are published and how many citations they each received. We measure the extent to which participation in TRECVideo evaluation has facilitated research which gets through the first hurdle of peer review to get published and secondly, in terms of bibliometrics, gets through the second hurdle of peer review and receives citations. Participating in TRECVideo is a means to an end, the ability to evaluate comprehensively one's research approaches, which in turn enable participants to build upon research and to convincingly disseminate their findings to the wider field of computer and information science. We can

have reasonable confidence that TRECVideo is ‘working’ if it makes possible a significant amount of research which is published and then cited. The purpose of this study is to examine how significant the research that TRECVideo has ‘made possible’ is through an analysis of the publications which have resulted from TRECVideo. We do this by examining a number of key questions about the publications and citations arising from TRECVideo: how many publications result from TRECVideo; what are they about; how often are they cited; what are the most popular venues in terms of paper and citation numbers; how important are TRECVideo papers to the citation profile of participants; how is TRECVideo ‘success’ linked to citation ‘success’? We address these questions using bibliometric and visualisation techniques. The next section provides a short introduction to TRECVideo to provide a wider context to the study.

### **TRECVideo: What it is.**

Information retrieval (IR) research has always had as one of its essential components, the systematic and repeatable benchmarking of any new technique for automatic analysis, indexing, retrieval, summarization or other content-based operation. Since the earliest days of information retrieval research (Sparck-Jones, 1981) the pioneers of this field including Luhn, Maron and Kuhns, Salton, Cleverdon, Sparck Jones, van Rijsbergen, Robertson and those who have followed them, have always used evaluation on test collections of data as the way in which the value of new theories, models, and ideas, are determined.

Up to the end of the 1980s, access to such test collections of data was very limited. Publicly available datasets were small, narrow in scope and over-used and there were no generally available large scale collections of documents, queries, and relevance assessments. In 1991 the National Institute of Standards and Technology (NIST) in the US organized the first Text REtrieval Conference (TREC) with the aim of building such a large scale collection of documents, queries and relevance assessment and allowing uniform evaluation using that dataset using a set of common and shared metrics. The growth rate associated with TREC throughout the 1990s is testimony to its success with increasingly large datasets being made available to the research community and evaluation metrics stabilizing. The IR research community effectively unified around TREC and its tasks so much that TREC started to branch out in terms of the nature of tasks and the variety of (text) data on which benchmarking was taking place. These new data and task types were known as “tracks” and in 2001 TREC launched a new track on video retrieval. This had the usual TREC mode of operation whereby NIST acquired and distributed (video) data to signed-up participants, NIST formulated and distributed search topics which participants executed on the video data using the systems they developed and they then submitted their top-ranked video clips for pooling and manual assessment by NIST personnel. This was then used to calculate performance metrics and at the TREC conference these results and the techniques behind the systems were shared and discussed.

The video track in TREC grew rapidly and in 2003, TRECVideo separated from TREC and became an independent, standalone benchmarking conference. Over the following 7 years TRECVideo operated on a variety of video genres and a range of content-based tasks including automatic detection of video shot boundaries, detection of semantic concepts within shots, fully automatic, semi-automatic and interactive search for video

shots or for known videos, near-duplicate video detection, video summarization, semantic event detection in CCTV and TV news story segmentation. All of these tasks are done in a hugely collaborative and supportive environment with sharing and donation of data and other resources among participants being the default, all in the name of progressing the field of video retrieval.

In this paper we focus on TRECVID during the year 2003 to 2009 inclusive. For the first 4 of those years the video data used was broadcast TV news, initially in the English language but then in 2005 and 2006 also including TV news in Chinese and Arabic. The video was accompanied by speech transcripts derived from automatic speech recognition which was automatically translated into English in the case of Chinese and Arabic. In 2007 TRECVID introduced a new genre of video provided by the Netherlands Institute for Sound and Vision, which consisted of general TV magazine programs. Also in 2007 TRECVID introduced camera rushes video, the raw, unedited video captured by a camera during the recording, and rehearsal of TV shows, provided by BBC. In 2008 TRECVID introduced CCTV camera footage taken in a major international airport.

The task of shot boundary detection ran from 2003 to 2007 at which point progress in the techniques seemed to have reached a plateau. The search task – automatic, manual and interactive – was introduced from the start and continues each year, as does the task of automatically detecting the presence of a set of semantic concepts. Automatic detection of TV news story bounds ran in 2003 and 2004, automatic summarization of BBC rushes video ran in 2007 and 2008, detection of events from surveillance data ran in 2008 and 2009 as did automatic detection of near-duplicate videos.

Participation by research groups in TRECVID increased every year except 2009, peaking in 2008 with nearly 80 groups and dropping to just over 60 in 2009. Participants come from all across the globe, and there is a great geographical spread. Some of the participants are regular and have taken part each year while others have just taken part once or twice; some participants represent larger research groups while others may be just a single PhD student working on a related topic. All, however, take part in the benchmarking in order to test out some new idea or technique they have developed. Each participant is allowed to submit more than one “run” for a task, including search, and participants usually vary some attribute or parameter of their search systems for each of the runs they submit.

## **Method**

A bibliometric study examining both the number of TRECVID publications and the number of times they have been cited gives us both quantitative and qualitative indicators of its scholarly impact. Citations counts are only one indicator of quality as studies show a variety of citing ‘motivations’, for a comprehensive review see Bornmann and Daniel (2008), but they do give an indication of the extent to which a publication has ‘made a difference’. Recently, there has been a growing recognition of how various data sources and citation metrics may impact on different disciplines and, in particular, the extent to which some established bibliometric tools may disadvantage computer science (Moed & Visser, 2007; Bar-Ilan, 2009). A considerable problem is one of coverage as any bibliometric tool can only accurately measure citations if it has data on all the

possible sources of publications and citations in any given discipline. Computer science publications are often conference or technical reports which are not comprehensively included in many of the standard citation analysis tools. Harzing (2010) analyses three different sources for citation analysis, ISI Web of Science, Scopus and Google Scholar across academics working in the Sciences and Social Sciences and Humanities. Computer Science is one of disciplines investigated and over a four year period she found that Computer Science was different from most sciences as Google Scholar provided five times as many citations as ISI. Another recent study by Freyne *et al.* (2010) also examined the citation scores from Google Scholar and ISI Web of Science for publications in computer science and found that lack of conference coverage by ISI put computer scientists at a disadvantage in evaluations based on ISI.

The table below provides a first cut of our search results using Scopus More and Google Scholar. In 2008, for example, Scopus More yielded 130 documents published as a result of TRECVID activity, Google Scholar for the same year has 586.

<u>YEAR</u>	<u>No of Publications Scopus General</u>	<u>No of Publications Scopus More</u>	<u>No of Citation Documents</u>	<u>No of Citations</u>	<u>H-Index</u>	<u>No of Publications Google Scholar (Using PoP)</u>	<u>No of Citations</u>	<u>H-Index</u>
2010	26	33	22	62	1	97	25	2
2009	85	40	62	56	4	401	590	11
2008	<b>130</b>	154	112	114	6	<b>586</b>	2516	19
2007	126	181	195	335	8	411	3655	28
2006	71	<b>241</b>	<b>263</b>	524	11	332	<b>3784</b>	<b>31</b>
2005	34	192	253	998	<b>15</b>	212	2497	28
2004	39	166	214	848	14	171	2195	23
2003	1	145	191	<b>1010</b>	14	58	1180	16
2002	0	3	3	3	2	9	52	4

TABLE 1: Initial pilot results showing comparison of Scopus and Google Scholar.

After consultation with experienced practitioners of citation analysis within computer science we decided, based on the issue of coverage, to use Google Scholar as our main source. ‘Publish or Perish’ (PoP)<sup>1</sup>, a software wrapper for Google Scholar, was used to manipulate Google Scholar searches. Whilst Google Scholar has the coverage we needed, its tools for checking duplicates and its ability to deal with large data sets, are not as advanced as its more established alternatives. The search results were therefore checked and cleaned manually. For each year we checked the PoP search results with the criterion for inclusion being ‘was this publication a direct result of TRECVID activity?’. By this we mean that the paper uses TRECVID data or benchmarking criteria or describes a technique tried in TRECVID. We excluded papers which just cited or mentioned TRECVID as our aim was to include papers truly derived from TRECVID participation. This cleaning eliminated most duplicates and also papers which were only tangentially related to TRECVID but which had been retrieved in the original literature search. The precision of the data set is, therefore, reasonably reliable but accurately checking the

<sup>1</sup> <http://www.harzing.com>

recall is more difficult and, despite our broad search strategies, some papers will have been missed. In terms of the results this means it is most likely that this study slightly under-reports the extent and impact of TRECVID publications.

Citation analysis is more commonly used for a single or a small group of authors and many of the tools used are designed towards this type of analysis. The TRECVID analysis, however, encompasses multiple authors, institutions, and publication types. The expert checking provided a good level of confidence that the variety of publications retrieved and analysed for this study were, in fact, about TRECVID.

## **The data set of publications**

Our main focus for this study was an investigation into the number and impact of publications written as a result of, or relying upon data from, the TRECVID conferences 2003-2009. The data set ‘TRECVID derived papers’ includes both the TRECVID conference papers known as workshop papers and published online on the TRECVID website, and papers published in different venues but based on TRECVID work. TRECVID publishes workshop papers describing how well the research groups’ techniques did against the evaluation process. These are not refereed, and *most* participants produce a paper, but unfortunately not all. Our initial pilots showed that TRECVID notebook papers were often highly cited which would suggest that they are used to support papers which have been published in other venues. This gives us an indication that they are having a scholarly impact. Thus, to gain an overall picture of the scholarly impact of TRECVID, it made sense to include them with the ‘TRECVID derived publications’ of which they consist of, on average, approx 15% of the total. In the first year of TRECVID, they consist of a much greater percentage of approximately 50%, but as the conference matured, more papers were generated for other venues, see Table 2.

After each annual TRECVID conference, many participants publish more detailed descriptions, or further experiments, or comparisons, or overviews, elsewhere, in journals, conferences or workshops. These will in nearly all cases have gone through a competitive peer review process to get published. So we examine all publications that exist because of the TRECVID conferences, either directly as a workshop paper, or indirectly but which could not have been written without the use of TRECVID data in some way. We then examine the citation patterns of these publications. The transition from non-referred workshop paper, to peer-reviewed publication to receiving citations we envisage as different ‘hurdles’ or stages of potential scholarly impact. Peer-review at publication is one stage of impact and citation is the second stage which shows us that a broader set of peers have also acknowledged and used the work. Thus publication and citation counts give us some indication of the overall impact of TRECVID related research.

We used this publication and citation data to investigate a number of questions, as outlined earlier, and these are discussed in more detail in the next section. The overall aim of these questions is to provide insights into the quantity and quality of TRECVID’s contribution to video retrieval research. We also present some initial investigations into possible factors that may influence the citation rates of different publications to see if these can inform our understandings of bibliometrics as a measurement tool for scientific ‘quality’.

## Key questions and answers

This section consists of a series of sub-sections, each examining one of the questions we raise about TRECVID, and providing analysis and answers.

### How many TRECVID publications are there?

Our study shows that for 2003-9 there were a total of 2,073 TRECVID-derived publications of which 310 were TRECVID workshop papers. As can be seen from the table below, as the conference matures, more publications reach venues outside the conference itself.

<u>YEAR</u>	<u>No TRECVID papers at CONFERENCE</u>	<u>No. TRECVID derived publication</u>	<u>No of citations</u>	<u>Cites per paper</u>	<u>H-Index</u>	<u>G-Index</u>
<b>2003</b>	28	64	1,066	<b>16.66</b>	18	<b>30</b>
<b>2004</b>	30	158	2,124	13.44	24	<b>40</b>
<b>2005</b>	37	225	2,537	11.28	28	<b>41</b>
<b>2006</b>	50	361	<b>4,068</b>	11.27	<b>30</b>	<b>52</b>
<b>2007</b>	48	382	3,562	8.97	28	<b>45</b>
<b>2008</b>	<b>64</b>	<b>509</b>	1,691	3.32	16	<b>23</b>
<b>2009</b>	53	374	780	2.09	12	<b>20</b>
<b>Totals</b>	<b>310</b>	<b>2,073</b>	<b>15,828</b>			

TABLE 2: Overview of data 2003-2009.

There is a steady overall increase in publication outputs, which is in line with the increase in participation. The year 2008 produced a particularly high number of publications as is shown in chart of publication trends below. This coincides with the year of greatest participation in TRECVID though one would expect at least a 1-year time lag with the publications from a TRECVID year following at least one year later.

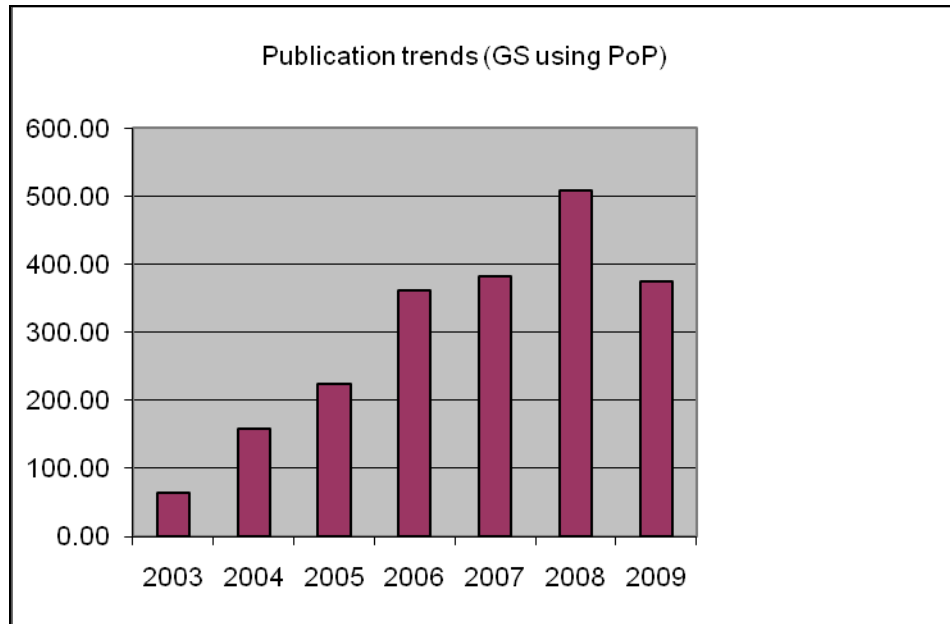


FIG 1: Overview of publication trends (Google Scholar using PoP)

### What are TRECVID papers about?

We also used the publication data to analyse and to visualize how the topics treated in TRECVID papers have developed and evolved year-on-year since the start of TRECVID. Similar analysis of topic development using tri-occurrence mapping, in IR as a discipline has been done by Sugimoto & McCain (2010). We examined the titles of all TRECVID-derived papers, the titles and abstracts of the most highly cited papers, and then the titles and abstracts of all TRECVID workshop papers.

#### *Using titles of all TRECVID-derived papers*

Using the titles of all 2,073 TRECVID-related papers in conferences, journals and workshops we generated word clouds for each year and compared between the years. This helps us analyse how popular sub-topics in TRECVID activities come and go each year and sometimes re-emerge in the later years.

Year	Most frequently used terms that year	Tasks exercised that year
2003	Shot, Segmentation, Boundary, Features, Framework, Transcript, Browsing	<ol style="list-style-type: none"> <li>1. Shot boundary determination</li> <li>2. News story segmentation</li> <li>3. High-level feature extraction</li> <li>4. Search</li> </ol>
2004	News, Segmentation, News, Semantic, Interactive, Story, Features	<ol style="list-style-type: none"> <li>1. Shot boundary determination</li> <li>2. News story segmentation</li> <li>3. High-level feature extraction</li> <li>4. Search</li> </ol>
2005	Semantic, Extraction, News, Concept, Annotation, Classification, Learning	<ol style="list-style-type: none"> <li>1. Shot boundary determination</li> <li>2. Low-level feature extraction</li> <li>3. High-level feature extraction</li> <li>4. Search</li> <li>5. Rushes exploitation</li> </ol>
2006	Semantic, News, Learning, Annotation,	<ol style="list-style-type: none"> <li>1. Shot boundary determination</li> </ol>



	Segmentation, Concept	<ol style="list-style-type: none"> <li>2. High-level feature extraction</li> <li>3. Search</li> <li>4. Rushes exploitation</li> </ol>
2007	Semantic, Learning, Concept, Annotation, Classification, News	<ol style="list-style-type: none"> <li>1. Shot boundary determination</li> <li>2. High-level feature extraction</li> <li>3. Search</li> </ol>
2008	Semantic, Annotation, Concept, Summarization, Learning, Rushes, Event	<ol style="list-style-type: none"> <li>1. Surveillance event detection</li> <li>2. High-level feature extraction</li> <li>3. Search</li> <li>4. Rushes summarization</li> <li>5. Content-based copy detection</li> </ol>
2009	Semantic, Concept, Annotation, Classification, Segmentation, Adaptive	<ol style="list-style-type: none"> <li>1. Surveillance event detection</li> <li>2. High-level feature extraction</li> <li>3. Search</li> <li>4. Content-based copy detection</li> </ol>

TABLE 3: Most frequently used terms in titles 2003-2009

Firstly, Table 3 shows how in general the topics of “Shot”, “Boundary”, “Segmentation”, “Features” (which have been the main research issues and interests since the research in video information retrieval took off in mid 1990s) were replaced with topics such as “Semantic”, “Concept” and “Learning” over the years. This represents the TRECVID community shifting interest, and incidentally the maturing of the field from a low-level, feature-oriented exploration to a high-level, semantic-oriented one. This change of dominant topics over the years is also in some way steered and guided by the “tasks” introduced in TRECVID each year as seen in the above table.

Using only paper titles might have been limited in terms of finding popular or important terms for specific approaches or techniques used each year as titles tend to pertain only high-level topic-related terms rather than more detailed, technical terms. Acknowledging this, we then extracted full titles and abstracts from the top 10 most frequently-cited papers to see if there are any obvious trends visible (as extracting all abstracts or full-text from all the papers would have been impractical). The result shows that while more technical terms indicating specific approaches or angles appear in the most frequent terms list (e.g. “classifiers”, “SVM”, “categorization”, “tags”, “speech”, “texture” and “edge”), there was no obvious change of usage frequency of these terms over the years. By using 10 most frequently cited papers of each year, we were capturing terms that were not only most used each year but propagated into the past years as “cited” means cited in the years subsequent to publication.

#### *Titles and abstracts of all TRECVID workshop papers*

Thirdly, we were interested in finding out frequency of terms used in each year’s TRECVID workshop papers to see what topics and methods were mentioned most frequently as a rough indication of their popularity or perceived interests in that year. We simply counted words frequency after removing stop-words such as “and”, “of” and “the”, as well as our own “TRECVID stop-words” such as “TRECVID”, “video”, “retrieval”, “search”, “baseline”, “experiment” and “digital”.

We performed term frequency analysis on the titles and abstracts from all 310 TRECVID workshop papers, in order to capture more technical terms that appeared in abstracts and also to represent those that appeared purely within the context of TRECVID

participation. Popular aspects or approaches in each year, can, in general be glimpsed in the term analysis with TRECVID workshop papers. Using this result, we generated a simple word cloud visualization taking top 35-40 terms for each year with Wordle<sup>2</sup>, and put each year's cloud next to each other to visually inspect the rising and falling popularities of terms (see Figure 2).

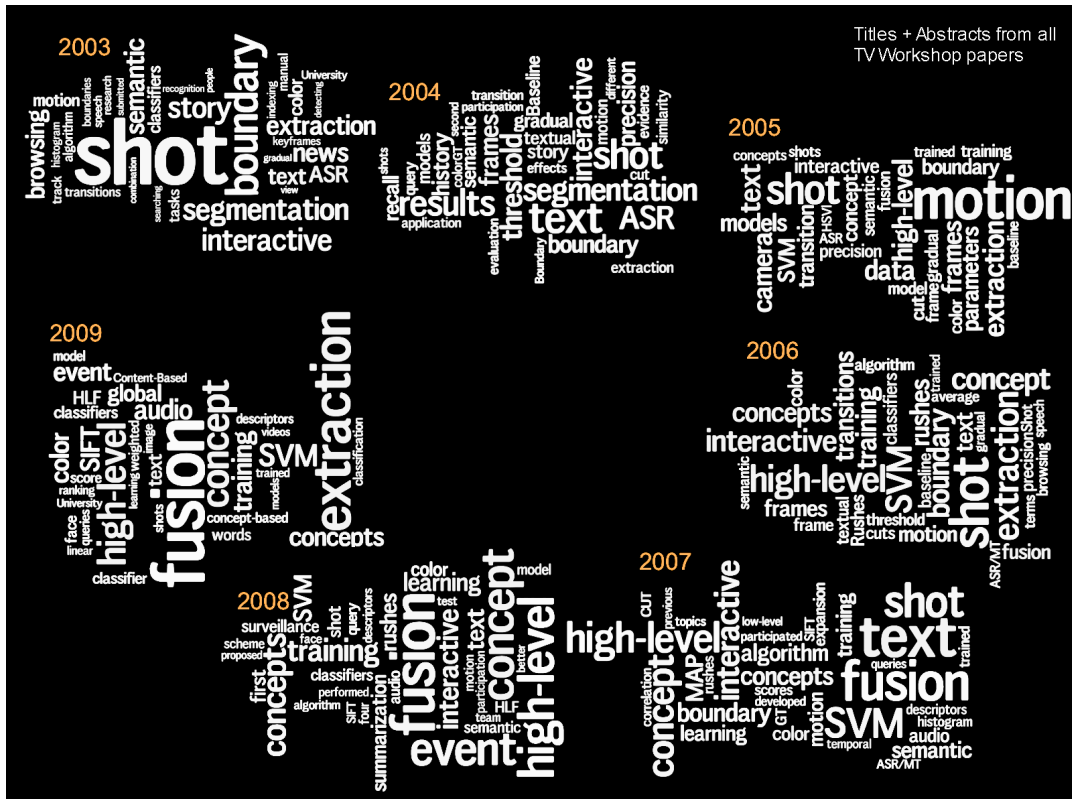


FIG 2: Wordle Visualization of top 35-40 terms used in titles and abstracts from TRECVID workshop papers 2003-2009.

As can be seen in Figure 2, low-level technical terms such as “shot”, “boundary”, “colour”, “motion” become smaller as years go on. Also notable in this visualization are the terms “concept” that appeared from 2005 and growing larger each year; the term “SVM” (Support Vector Machine) that appeared from 2005 and grows larger and more or less staying on thereafter; the term “fusion” that started in 2006 and became very popular subsequently; the term “SIFT” (Scale-Invariant Feature Transform) that appeared in 2007 and grows bigger and bigger each year; terms “ASR” (Automatic Speech Recognition) and “text” appear throughout the years indicating that the use of non-visual cues to help video retrieval has been attempted throughout the TRECVID activities.

Finally, we created a bead diagram representing the top 20 terms appearing across the 7 years of TRECVID where the font size of the word represents the importance of that word in that year. This is shown in Figure 3.

<sup>2</sup> <http://www.wordle.net>

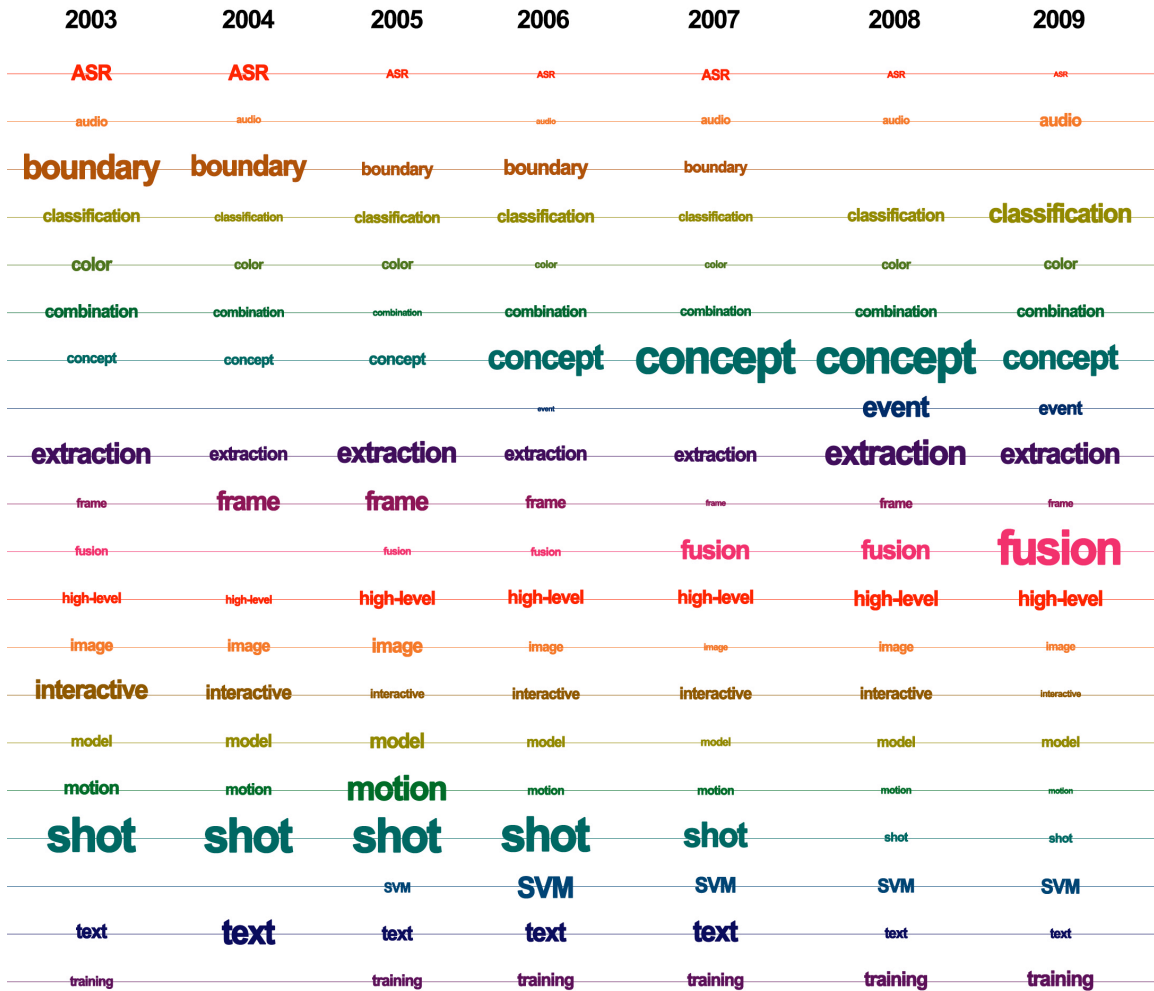


FIG 3: Bead plot of the most important 20 words across the 7 years of TRECVID.

Similarly to the trend seen in Figure 2, the bead plot in Figure 3 also shows the diminishing of some topics (“shot”, “boundary”, “ASR”), the growth of others (“concept”, “fusion”, “high-level”, “SVM”, “training”) while other topics remain fairly constant. This is in line with expectations and corresponds to the ending of tasks (shot boundary, use of ASR in search), and the emergence of new techniques for high-level concept detection based on training support vector machines (SVM).

### How often are TRECVID papers cited?

The total number of citations over the time period of the conference was 15,828 and the average cite per paper was 9.58. The citation rates of TRECVID papers are skewed in that a small number of papers receive a very large number of citations with this quickly tailing off. This shows that the citation patterns of TRECVID papers conform to a distribution pattern which has often been observed in other bibliometric studies (Price, 1976). The charts below show citation distribution rates for the year 2007 and also citation trends between 2003-2009. Note that 2007 is chosen as a representative year as it

is still relatively recent whilst not being so recent that it is likely to accrue many more citations than it already has.

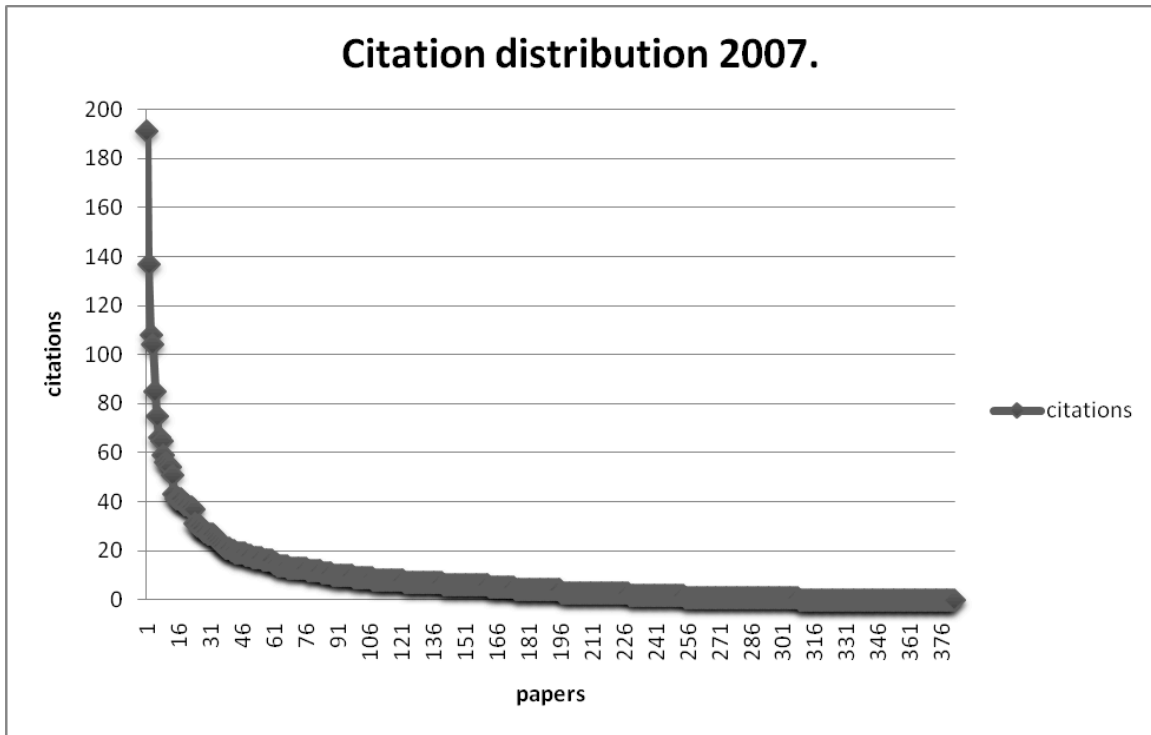


FIG 4: Citation distribution 2007.

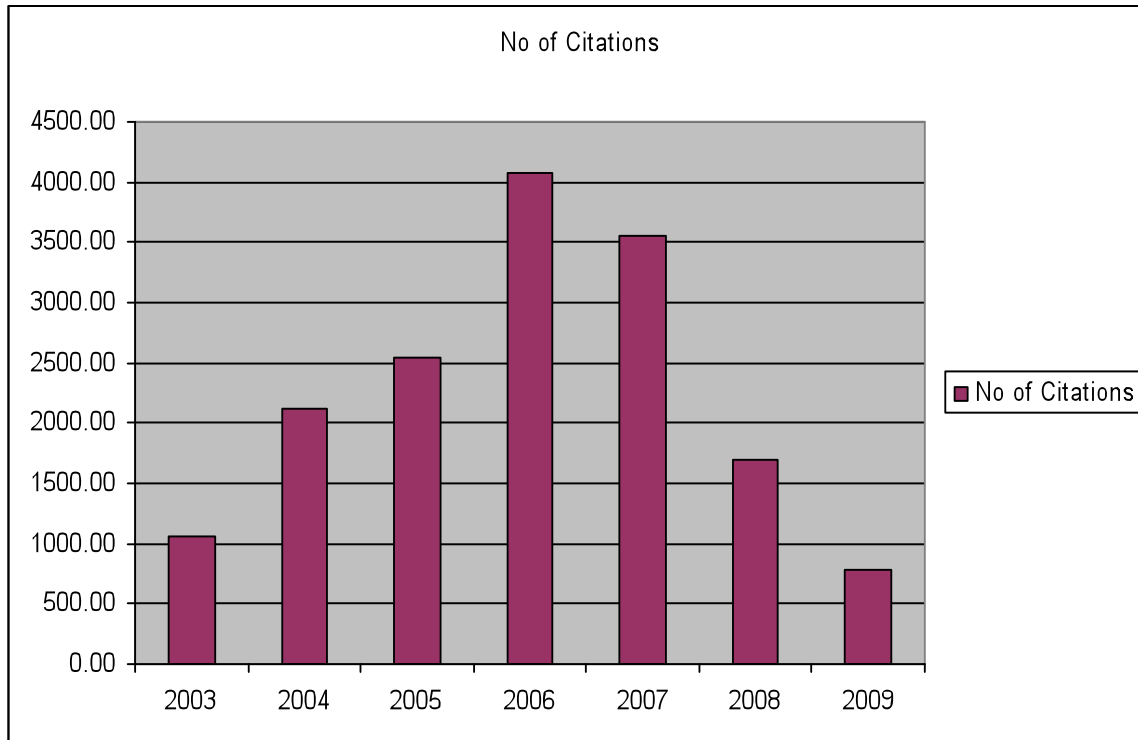


FIG 5: Citation trends 2003-2009.

The TRECVID-derived publications in 2006 have, so far, the largest number of citations. One of the reasons for this is that there is a recommended citation suggested to participants for when TRECVID is referenced in scholarly publications, and the 2006 recommended citation received over 400 citations. This paper skews the peak in 2006 somewhat and if that was removed then the distribution of citations across the years would be more even.

Our next two questions investigate some possible reasons why some papers are cited much more than others. Firstly, we investigate whether success at TRECVID in terms of system performance against the evaluation criteria, leads to success in terms of citations. Secondly, we examine where highly cited papers are published to see if any particular venues appear to attract more citations. In both cases, particularly the latter, it is problematic to assert causation due to the multiple other factors that can influence citation, but some patterns can be observed.

### **Does 'good performance' in TRECVID lead to 'high performance' in citations?**

Do teams who develop techniques which 'score' highly in TRECVID then go on to produce papers which then also 'score' highly in terms of the number of cites they get, or to put it another way, do people tend to cite papers that describe techniques which were successful at TRECVID more than papers which describe less successful techniques?

We investigated this question by firstly identifying the top performers and the lower performers at TRECVID 2006 using the criteria discussed below. We then analyzed the citation rates of TRECVID derived papers published in 2007 written by those team members, with the assumption that these would have been mainly about work done in 2006.

TRECVID 2006 was the final year of the cycle of using broadcast TV news as the video source before moving on to use video data from the Netherlands Institute of Sound and Vision. The tasks in 2006 were shot boundary detection (SBD), feature detection, and search, the latter two based on a master shot reference supplied by the organizers. A rushes video summarization task was also on offer as an exploratory task but few groups completed this and there was no formal scoring or feedback to participants in this task. The video used for the SBD, feature and search tasks consisted of 159 hours of TV news from November/December 2005, with news being from TV stations speaking English, Chinese and Arabic, with most of the data being Arabic. Output from an Automatic Speech Recognition (ASR) system was provided for the video, with machine translation into English for the Chinese and Arabic. All this meant that the quality of the text (from ASR or from ASR followed by machine translation) was quite poor in terms of accuracy, forcing participants to focus on visual aspects of content-based retrieval.

In addition to the master shot reference, the MediaMill group at the University of Amsterdam provided the output of 101 automatic feature detectors on the search data, and a group from Columbia University, Carnegie Mellon University and IBM provided the output of manual annotation by 449 features from the LSCOM (large scale concept ontology for multimedia) ontology also on the search data, for all participants to use.

In 2006, 54 participating groups completed one or more of the tasks, broken into 26 who completed SBD, 30 who completed feature detection and 26 who completed at least one form of the search task. Many of these groups went on to publish further details on their TRECVID 2006 work elsewhere, but determining which were the best-performing groups in order to correlate that with subsequent publication and citation is difficult because not all groups did all tasks and even for those who did, they may have performed better in some tasks than in others. This means that a ranking of groups taking part would not only be against the spirit of participation in TRECVID but would also be impossible.

Instead, we have selected from among the 2006 participants, two clusters of three participants each, all of whom have taken part in both the feature detection and the search tasks. We rationale this on the basis that these are the most difficult of the tasks and groups who have completed both these make a serious and large commitment to participation. The first group of three all score highly in each task, certainly within the top-5 or top-6 in each task whereas the second group consists of teams who are not the top-ranked in either but have mid-range performances.

The results of our analysis show a strong connection between high performance in TRECVID and high performance in citations. This was not the only factor in high citation counts, for example, as observed in other studies (Asknes, 2006), review papers were often in the top-cited papers of each year. The key findings are that TRECVID ‘top scorers’ do nearly twice as well than average in their citation scores, and three times as well as ‘low scorers’. TRECVID ‘top scorers’, however, do not do as well in citation count as some other papers written by ‘non top scorer’. In the top 25 most cited papers ‘non top scorer’ papers do better than ‘top-scorers. Being a ‘top scorer’ is a good indicator of citation success as is shown in table below for cites per paper comparison of high performing TRECVID team versus the average cites per paper for the entire year. Further detailed analysis of all the top scoring papers, as done recently for ACM

published papers by Wainer, de Oliveira & Anido (2010) would provide further insights on this questions but was beyond the current scope of this study.

The table and charts below shows the publication and citation impact for 2007 (based on research done in TRECVID 2006) broken down by ‘all papers’, ‘low scoring teams’ and ‘top scoring teams’.

<b>Breakdown of figures for research teams in 2007.</b>			
	<b>Cite count</b>	<b>Paper count</b>	<b>Mean cite per paper</b>
<b>3 top scoring research teams</b>	990	55	18.0
<b>3 low scoring research teams</b>	48	8	6.0
<b>Other papers</b>	2524	319	7.9
<b>Overall in 2007</b>	3562	382	9.3
<b>Overall without top scoring</b>	2565	327	7.84
<b>Papers of other research teams in top 25 cited papers</b>	1011	15	67.4
<b>Top 25 cited papers overall</b>	1564	25	62.56
<b>Top 3 scoring research teams in the top 25 cited papers</b>	553	10	55.3

TABLE 4: Research teams 2007.

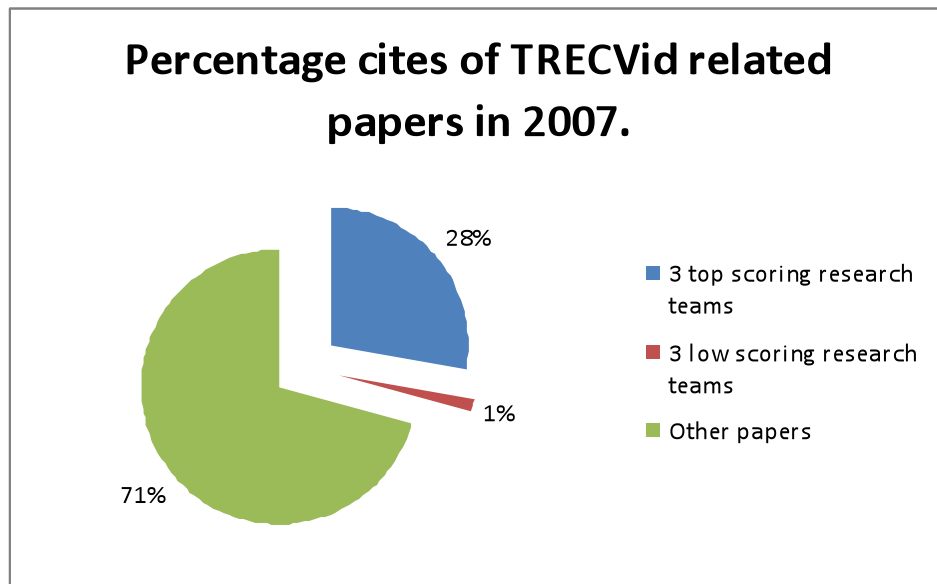


FIG 6: Percentage of citations of TRECVID related papers 2007.

### Percentage breakdown of TRECVID related papers in 2007.

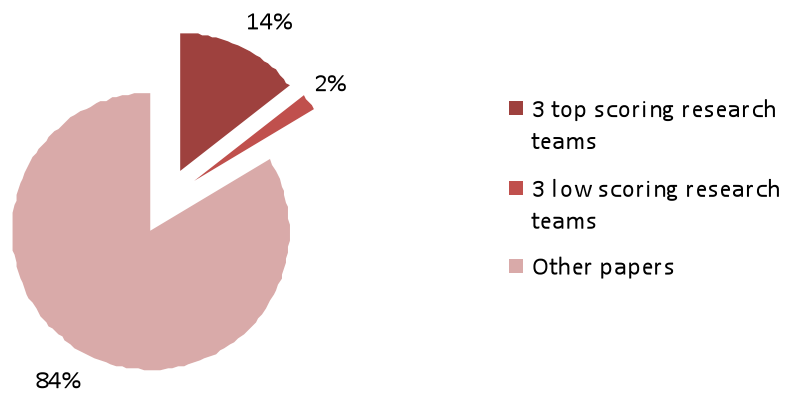


FIG 7: Percentage breakdown of TRECVID related papers 2007.



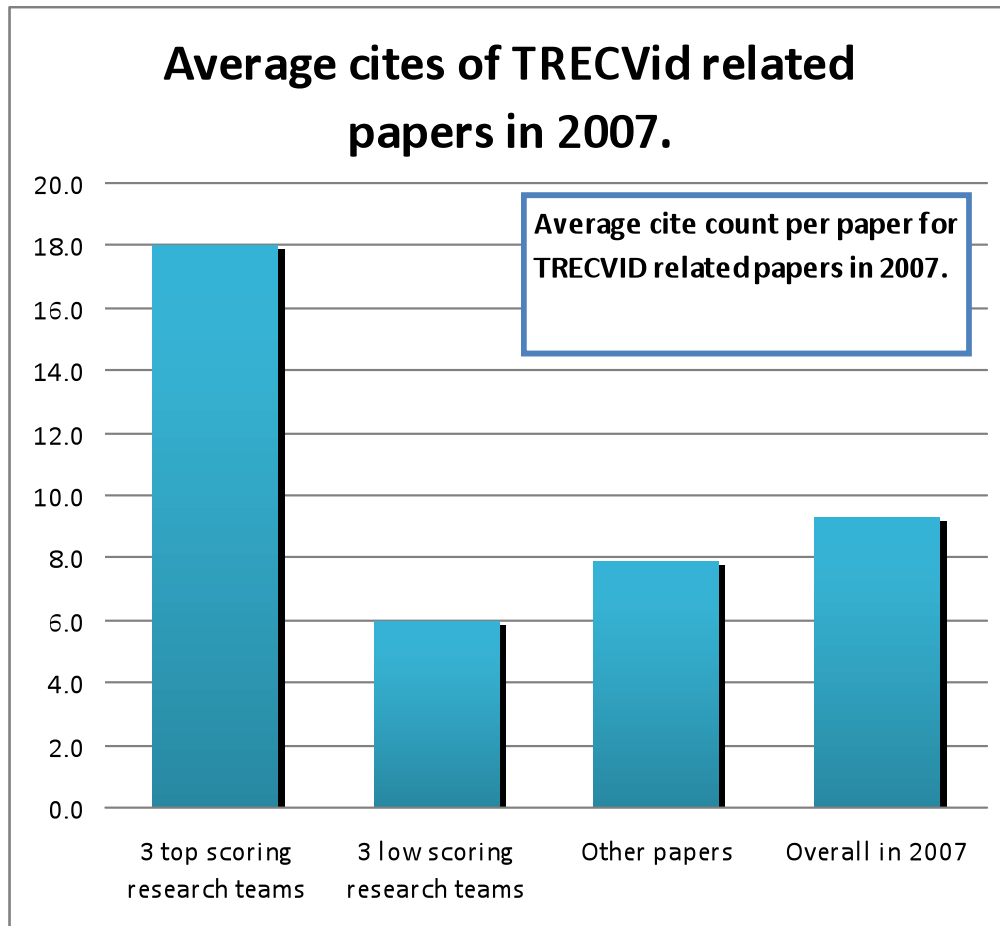


FIG 8: average cites of TRECVID related papers 2007

Teams who performed less well at TRECVID also did less well in terms of citations suggesting that papers discussing techniques which are not successful are not cited as much as papers discussing more successful techniques. This is, perhaps, not altogether unexpected, but it raises some interesting question about the relationships between citations, technological progress and science. Data on techniques which don't perform well at TRECVID still make an important contribution to progress by eliminating certain lines of development. In terms of retrieval performance they may be not be successful but, in terms of science, they form part of the progress. The goal is collective and it is in the best interests of the field if a variety of techniques are tried out some of which, inevitably, will do less well than others.

### Where are (highly cited) TRECVID papers published?

Here we examine the publication venues of TRECVID-derived papers to analysis both popular venues (which publish a high number of TRECVID-derived papers) and high impact venues (where TRECVID-derived papers attract a lot of citations). Conference and journal venues 2007-2009 were investigated. We split the publications data set by source, either conference or journal, and then ranked them by citation counts. This also provided an overview of the relative importance of journals and conferences in terms of the citation count of the TRECVID papers, see table and chart below. We see that, in line

with other bibliometric studies within computer science, conferences are significantly more important in terms of publications and citations than journals.

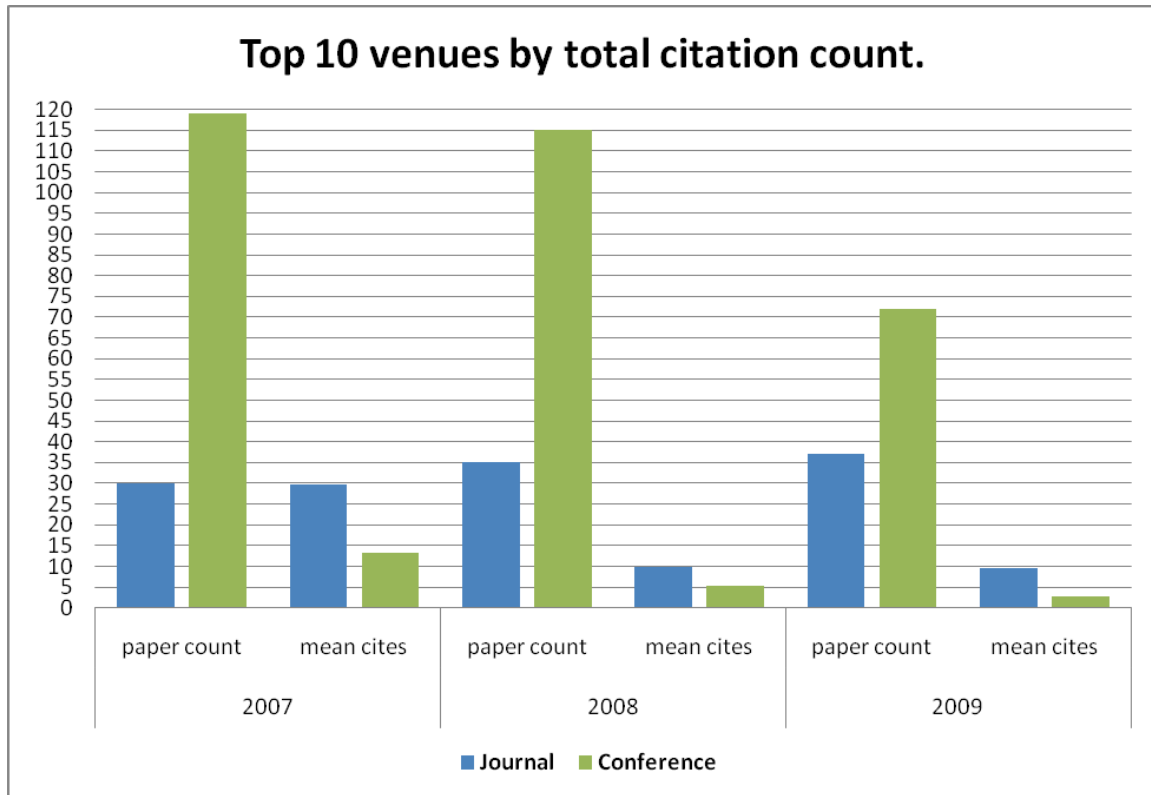


FIG 9: Paper counts and mean cites to TRECVID-derived papers in journals and conferences 2007 – 2009.

*Conferences 2007-2009*

Figure 10 shows the ranking of top 10 conference venues in 2007, 2008 and 2009, which had highest number of cites to TRECVID-derived papers in each year.

2007	2008	2009	
1. CIVR (565)	1. ACM MM (147)	1. ACM MM (51)	ACM CHI SIGCHI Conference on Human Factors in Computing Systems
2. ACM MM (497)	2. CIVR (130)	2. CIVR (45)	ACM MM ACM International Conference on Multimedia
3. MIR (238)	3. MIR (91)	3. CVPR (41)	BMVC British Machine Vision Conference
4. ICME (114)	4. ECCV (60)	4. ICCV (17)	CBMI International Workshop on Content-based Multimedia Indexing
5. SAMT (75)	5. ICME (47)	5. CLEF (15)	CIVR International Conference on Image and Video Retrieval
6. CVPR (42)	6. BMVC (33)	6. ICME (13)	CLEF Workshop on Cross-language Information Retrieval and Evaluation
7. KDD (17)	7. SIGIR (31)	7. ICML (9)	CVPR IEEE Conference on Computer Vision and Pattern Recognition
8. ACM CHI (16)	8. CVPR (28)	8. WWW (5)	ECCV European Conference and Computer Vision
9. SIGIR (13)	9. ICIP (20)	9. ICDMW (4)	ICCV International Conference on Computer Vision
10. ICCV (13)	10. CBMI (20)	10. WIAMIS (4)	ICDMW IEEE International Conference on Data Mining Workshops
			ICIP IEEE International Conference on Image Processing
			ICME IEEE International Conference on Multimedia and Expo
			ICML International Conference on Machine Learning
			KDD ACM International Conference on Knowledge Discovery and Data Mining
			MIR ACM International Conference on Multimedia Information Retrieval
			SAMT International Conference on Semantic and Digital Media Technologies
			SIGIR SIGIR Conference on Information Retrieval
			WIAMIS International Workshop on Image Analysis for
			WWW Multimedia Interactive Services
			International World Wide Web Conference

FIG 10: Top 10 conferences by year where TRECVID-derived papers was mostly frequently cited 2007-2009 (number in brackets shows the total number of citations to TRECVID-derived papers in the conference that year).

As the figure shows, between 2007 and 2009, TRECVID papers were consistently being cited especially at the high-profile multimedia/image processing venues such as the International Conference on Image and Video Retrieval (CIVR, rank 1 in 2007, then rank 2 in 2008 and 2009) and the ACM International Conference on Multimedia (ACM MM, rank 2 in 2007, then rank 1 in 2008 and 2009). This shows that they are getting through first hurdle of competitive peer review and then also receiving citations after publication. While hard-core image/video processing and computer vision conferences such as CVPR and ICCV are seen citing TRECVID papers during these three years, less image/video-centric events such as ACM CHI (International Conference on Human Factors in Computing Systems) and CLEF (Workshop on Cross-Language Information Retrieval and Evaluation) are also seen citing TRECVID papers indicating its impact spilling over to other neighbouring disciplines. It is, of course, difficult to know the relative influence of the ‘quality’ of the conference venue or the ‘quality’ of paper in terms of attracting citations and conference series will vary in their attraction from year to year depending on the venue, but we can confirm that TRECVID papers are appearing at a widespread set of venues.

#### *Journals 2007-2009*

Figure 11 shows which journals citing TRECVID-derived papers are published in.

2007	2008	2009
1. IEEE Trans. Multimedia (294)	1. IEEE Trans. Multimedia (100)	1. IEEE Trans. Pattern Analysis and Machine Intelligence (138)
2. Pattern Recognition (191)	2. Proc. the IEEE (72)	2. Computer vision and image understanding (55)
3. IEEE Trans. Circuits and Systems (108)	3. J. Visual Communication and Image Representation (35)	3. IEEE Trans. Multimedia (49)
4. ACM Trans. Multimedia Computing, Communications and Applications (104)	4. Computer vision and image understanding (32)	4. Foundations and Trends in Information Retrieval (26)
5. IEEE Trans. Pattern Analysis and Machine Intelligence (39)	5. IEEE Trans. Systems Man and Cybernetics (27)	5. Pattern Recognition Letters (22)
6. Information Systems (38)	6. Annual Review of Information Science and Technology (21)	6. IEEE Trans. Image Processing (18)
7. IEEE multimedia (37)	7. ACM Trans. Multimedia Computing, Communications, and Applications (16)	7. IEEE Trans. Circuits and Systems (11)
8. Image and Vision Computing (29)	8. IEEE Multimedia (16)	8. Machine Learning (11)
9. Signal Processing: Image Communication (26)	9. J. Electronic Imaging (12)	9. Multimedia tools and applications (10)
10. Multimedia Systems (22)	10. J. Information Science (11)	10. Pattern Recognition (9)

FIG 11: Top 10 journals by year where TRECVID-derived papers are most frequently cited 2007-2009 (number in brackets shows the total number of citations to TRECVID-derived papers in the journal that year).

An important journal for TRECVID papers is IEEE Transactions on Multimedia (rank 1 in 2007 and 2008, then rank 3 in 2009) as, apart from 2009, it is the journal which receives the highest total number of citations for all its TRECVID-related papers. In a similar way to the conferences we can see that TRECVID papers are being published in the top quality computer science journals covering the field. IEEE and ACM Transactions seem to be the most popular journals during these three years where TRECVID-derived papers are cited. Some journals only occur in some years, for example, in 2008 the Annual Review of Information Science and Technology and the Journal of Information Science (ARIST), traditionally more information science than computer science publications, are in the top ranking. ARIST 2008 had a paper on ‘Visual image retrieval’ by Peter Enser (2008a) which explains its ranking in that year and, likewise, the Journal of Information Science had a paper by the same author (Enser, 2008b) on ‘The evolution of image retrieval’. These were current ‘state of the art’ papers reviewing progress and some of their content discussed the role of TRECVID but they are clearly a different kind of paper than one by a participant describing new breakthroughs or techniques.

### Impact on Careers

We now look at the impact and influence that TRECVID has had on the careers of individuals by examining the publication and citation patterns of 5 typical TRECVID participants who range from early to late career stage. The total number of participants in TRECVID from 2003 to 2009 is 1,099 but here we select a sample of 5 in order to examine the role that TRECVID publications have played in their publication output and citation counts when compared to their non-TRECVID papers between 2003 and 2009.

The data is cumulative so we look at the relative influence of TRECVID papers on their citation scores as their career has progressed. We call these individuals tv1, tv2, tv3, tv4 and tv5 in ascending order of seniority. Figure 12 compares the cites per paper for TRECVID papers and non-TRECVID papers, among the five researchers over the five year period.

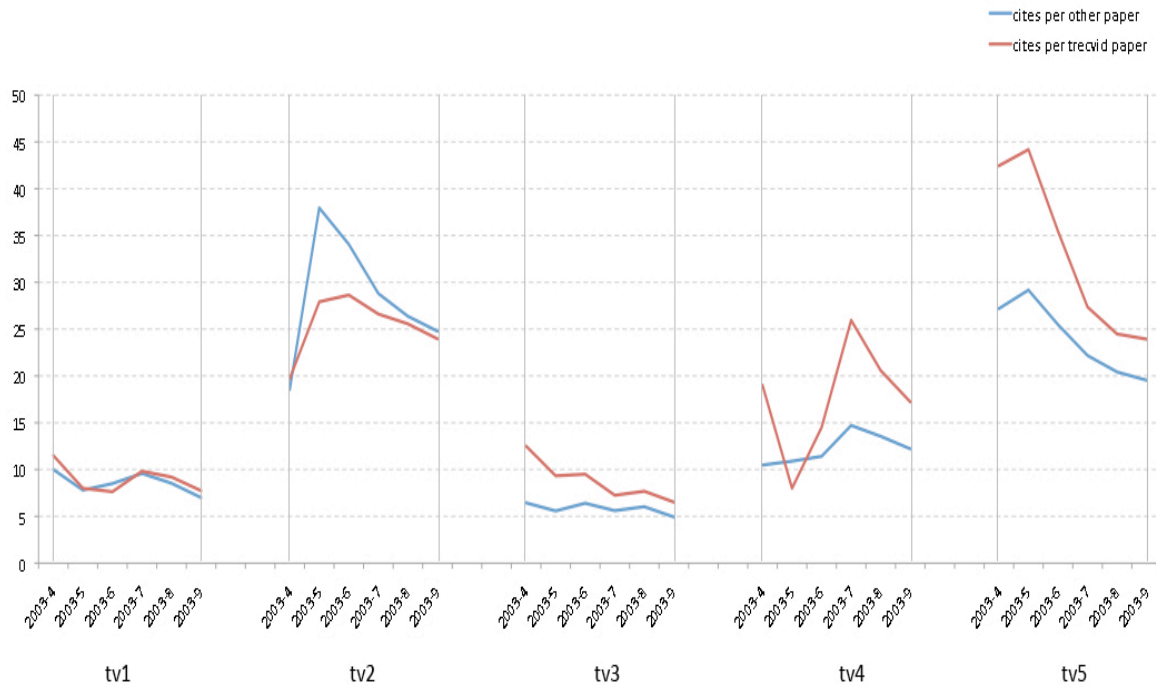


FIG 12: Number of cites to TRECVID papers vs. non-TRECVID paper among 5 different researchers in different stages of their career (most junior ‘tv1’ to most senior ‘tv5’).

In Figure 12, tv1 (most junior researcher) naturally has the least number of publications overall and tv5 (most senior researcher) has the highest number of publications overall, and the other three researchers (tv2, tv3 and tv4) are somewhere in between. However, across all five researchers we can see that TRECVID papers receive more citations per paper than their other papers. This trend seems more marked as the career progresses in the increasing gaps between TRECVID papers and non-TRECVID papers from tv1 to tv5. It is difficult to ‘separate’ TRECVID publications and non-TRECVID publications completely as clearly all a given researcher’s work is inter-related. If a paper isn’t about TRECVID for one of these participants, is it likely to be informed by it and *vice versa* for his or her TRECVID papers. This does show that, in most cases, for participants in TRECVID their TRECVID-related work receives more citations per paper than their other work. In terms of bibliometric measures, which are increasingly important in academic promotion and recruitment, this suggests that TRECVID-related work is a good use of their research time.

## 6. Discussion and conclusions

TRECVID has resulted in a large number of ‘spin off’ or derived publications which have received a substantial number of citations in total with some of them being

very highly cited. Research carried out at TRECVID has impacted on the field of video research through publication in high quality conference and journals venues and also through being cited by other researchers working in similar or related fields. We can see from the visualizations of TRECVID topics over time how new approaches have been developed through TRECVID. For those involved in TRECVID, their publications relating to the conference have made a significant contribution to their overall research impact. We cannot, of course, know what would have happened to these research ideas or researchers if TRECVID had not taken place, as this would require a control in which TRECVID had not happened.

What does this study tell us about bibliometrics and its reliability and validity as way of measuring scholarly impact? What does it tell us about what scholarly impact actually is? In terms of reliability it reinforces previous work already discussed about problems of coverage for computer science in the established bibliometric tools of Web of Science and Scopus. Publish or Perish, based on Google Scholar, has almost astonishingly better coverage. Despite this, due to the expert checking, we know that it missed some papers. Publish or Perish also has limitations to its 'ease of use' and functionality, particularly for large data sets, compared to its more established rivals. A more detailed paper on methodological issues, describing 'lessons learned' from our chosen methodology as a guide for future related studies, will be published elsewhere and the data used in our study (will be) available at (website to be set up).

In terms of the validity of bibliometrics in general, the main question is whether a high citation rate (quantitative) for a research paper actually tell us something about the quality of that research paper and, by extension, its authors and perhaps their department or institution? Our main contribution to this debate is the investigation into the relationship between TRECVID performance and citation performance. This data strongly suggests that 'success' at TRECVID does lead to 'success' in citation. Thus, one could argue, that citation counts do 'measure' quality if we accept that research quality is about finding solutions to problems that work better than other solutions proposed so far. The setup in TRECVID is, in one sense, a microcosm of science. In a very limited and finite world, researchers test hypotheses, or at least proposed approaches, against a data set. Some of these turn out to work well and some do not. For TRECVID and, more importantly, the wider field of video retrieval these 'failures', once confirmed as 'falsified' hypotheses (or more accurately proposed approaches) to use Popper's (1959) terminology, will be important in shaping the research and development trends of the future. Thus in using bibliometrics to measure quality we need to be clear that progress may rely on some researchers not doing too well and coming up against dead ends. They may, during this time, not receive many citations but they may, nevertheless, still make an important contribution. Our understanding of the relationship between citation rates and quality, in terms of what scholarly impact actually means, should include an awareness of this.

## **Acknowledgements**

This material is based upon work supported by Science Foundation Ireland under Grant No. 07/CE/I1147. Thanks to Julia Barrett of UCD Library and Shane McLoughlin of the UCD School of Information and Library Studies for invaluable assistance to this project.

## References

Asknes, D.W. (2006). Citation rates and perceptions of scientific contribution. *Journal of the American Society for Information Science and Technology*, 57 (2), 169-185.

Bar-Ilan, J. (2009). Which h-index?-a comparison ofWos, Scopus and Google Scholar. *Scientometrics*, 74 (2), 257-271.

Bornmann, L., & Daniel, H-D. (2008). What do citation counts measure? A review of studies on citing behaviour? *Journal of Documentation*, 64 (1), 45-80.

Enser,P.G.B. (2008a). Visual Image Retrieval, *Annual Review of Information Science and Technology*, 42.

Enser, P.G.B. (2008b).The evolution of visual image retrieval, *The Journal of Information Science*. 34 (4), 531-546

Freyne, J., Coyle, L., Smyth, B., Cunningham, P. (2010) Relative Status of Journal and Conference Publications in Computer Science. *Communications. ACM*. 53(11), 124-132.

Harzing, A-W. (2010).Citation analysis across disciplines: The impact of different data sources and citation metrics. Retrieved October 2010, from [http://www.harzing.com/data\\_metrics\\_comparison.htm](http://www.harzing.com/data_metrics_comparison.htm)

Moed, H.F.,& Visser, M.S. (2007). Developing bibliometric indicators of research performance in computer science: an exploratory study. Research report to the council for physical sciences of the Netherlands organisation for scientific research (NWO). CWTS Report 2007-01. Retrieved October 2010, from [http://ict.nwo.nl/files.nsf/pages/NWOA\\_78NJ63/\\$file/CWTS\\_Computer\\_Science\\_Study.pdf](http://ict.nwo.nl/files.nsf/pages/NWOA_78NJ63/$file/CWTS_Computer_Science_Study.pdf)

Popper's (1959). *The logic of scientific discovery*. London, UK: Hutchinson.

Price, D.D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27 (5), 292-306.

Robertson, S. (2008). On the history of evaluation in IR. *Journal of Information Science*, 34 (4), 439-456.

Rowe, B.R., Wood, D.W., Link, A.N. and Simoni, D.A. (2010) Economic Impact Assessment of NIST's Text REtrieval Conference (TREC) Program. RTI International, Project Number 0211875. Retrieved October 2010, from <http://trec.nist.gov/pubs/2010.economic.impact.pdf>

Sparck-Jones, K. (1981). *Information Retrieval Experiment*. London, UK: Butterworths.

Sugimoto, C.R., & McCain, K .W. (2010). Visualising Changes over time: A history of information retrieval through the lens of descriptor tri-occurrence mapping. *Journal of Information Science*, 36(4), 481-493.

Wainer, J., de Oliveira, H.P., Anido , R.(2010) . Patterns of bibliographic references in the ACM published papers. *Information Processing and Management*.  
Doi:10.1016/j.ipm.2010.07.002