A CREATIVE INTELLIGENT OBJECT CLASSIFICATION SYSTEM USING GOOGLE'S™ IMAGES IMPORT SEARCH FUNCTION

N.J. LUWES

Abstract

Limits of artificial intelligent, expert systems are defined by the specific hardware limitation of the specific system. Limits can be overcome, or addressed, by giving an intelligent system web access; therefore giving it access to Google's™ vast hardware, search functions and databases. Reverse image searches can be done directly in Google's™ image search bar since October 2011.This reverse image search function is used by the proposed system to do object recognition. Computational creativity, or the ability of a program or computer to show human-level creativity and interaction, is achieved by means of a voice communication of the object identification result to the user. The proposed system interprets the result by doing a definition web search and communicating this to the user. The results show that with the novel interpretation software, it should be possible to use Google™ as an artificial intelligent, computational creative system.

This proposed system thus has the ability to do object classification by accessing Google's™ vast hardware, search functions and databases, thereafter would the proposed system search a suitable definition for the classification. All of this information is communicated to the user using voice.

These techniques could be used on an automatic guided vehicle, robots or expert systems.

Keywords: creative intelligences, image processing

1. INTRODUCTION

Computer vision can be defined as the process for acquiring, processing, analysing, and understanding images. Computer vision applications range from industrial machine vision systems to artificial intelligence and computers or robots that can comprehend the world around them (Szeliski, 2011, p.5; Miller et al,2011, p.1). Object classification is an important aspect of this comprehension.

An ideal artificial intelligent system would have creative computation ability. Computational creativity is the ability of a program or computer to show human-level creativity and interaction (Pereira, 2007, p.205-p.207; Duch, 2006, p.435). Creative computation ability is achieved though human like interaction e.g. a computer or program mapping a room labeling the objects with serial numbers. As humans we want it rather to be a chair or a table, not object 123_4 or object 9998_0.

A creative intelligent system would be able to explain the function of the object to a human with human like communication or protocol (using voice to communicate to the users).

Limits of artificial intelligent or expert systems are defined by the specific hardware limitation of that system (Slack, 1991, p.117). In other words to improve the capability of an artificial intelligent or expert system, one would need larger storage space for training and operational data. Furthermore fast and large hardware processors will be required. Real time object classification uses complex algorithms that are hardware intensive and time consuming. A solution would be to use remote hardware and storage. By giving the intelligent system internet capability one can access the vast processing power of Google's™ hardware as well as Google's™ image data base and images processing algorithms.

Google's™ hardware capability is kept secret for competitive reasons, but an estimation of its hardware was made by Jonathan G. Koomey, Ph.D. Consulting Professor, Stanford University. He published data center electricity use from 2005 to 2010. He calculated that the total worldwide use in 2010 was 198.8 billion kWh. with an estimated 900,000 servers (Koomey, 2011, p.3,10,13). This hardware is indirectly accessible when using Google™.

Artificial intelligent ability corresponds to the size of its training data sets. The quantity of images on Google™ is stated in 2001 to have been 250 million images and by 2010 it reached 10 billion images indexed (Google™ official blog).

Google™ images were first introduced in July 2001. It allows users to search the web for image content (Google™ official blog) was and is keyword driven. Since October 2011 it has been possible to do reverse image searches directly in the image search-bar. This reverse image searches function is used by the proposed system.

Creative intelligence is achieved by using the results used as a Google™ search keyword and the expert system would communicate an explanation of the result using the vast data base of the web. A philosophical aspect of this system is that the information uploaded on the web is fed by individuals giving this and equivalent systems a human trained creativity.

Evaluation is done using three chairs as objects in the experiment.

A. B. C.

Fig. 1. Objects A, B and C

All three samples are classified as a chair, although they are visually different. Furthermore no data base of images on the local hardware exists. Therefore all objects shown to the system would be the first time the system would “visualise” the objects.

The system is designed to do preprocessing, optimising the image for a web search. Optimal results will be achieved if a descriptor is added where the object was “seen”. The system firstly reads through the web results from Google™ producing the most likely answer. Secondly the result is communicated to the user via voice and on screen display. The next step in creative computing is that the result will be added as a web search on the online dictionaries. The system would read through the web results producing the most likely definition or explanation and communicate this to the user.

2. METHODS

The system is divided up into three parts namely:

- The picture acquisition and image processing.
- The image search function and result voice communicator
- The result search function and voice communicator

3. PICTURE ACQUISITION AND IMAGE PROCESSING

Images were taken with a Sony alpha 200 at a resolution of 10.2 mega pixels with a 70-18 lens, an aperture of 4.5 and ISO 100. For evaluation the images of each object were taken, rotating it with 45 degree angles (generating samples at 0, 45, 90, 135, 180, 225, 270 and 315 degrees).

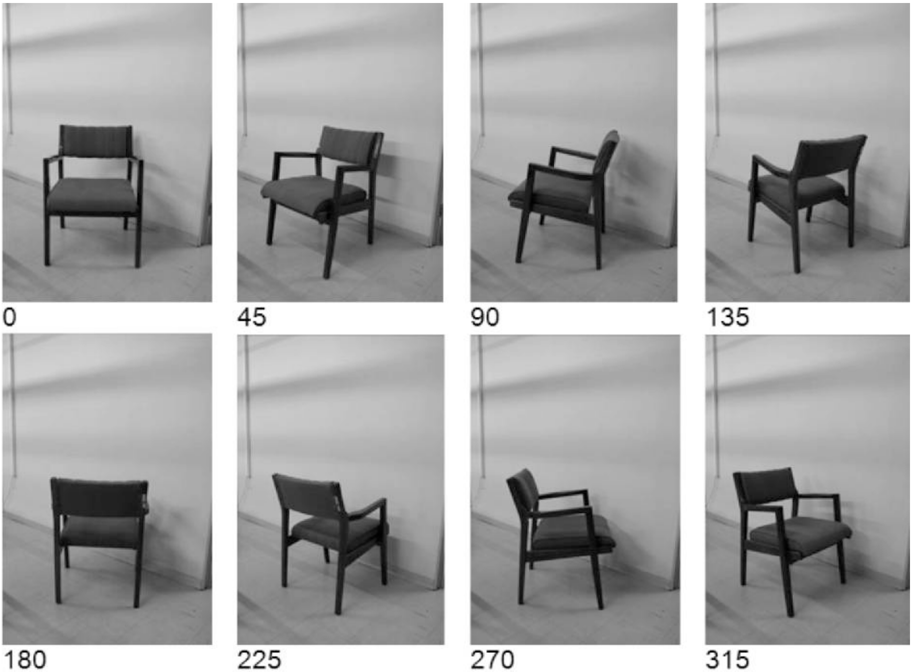
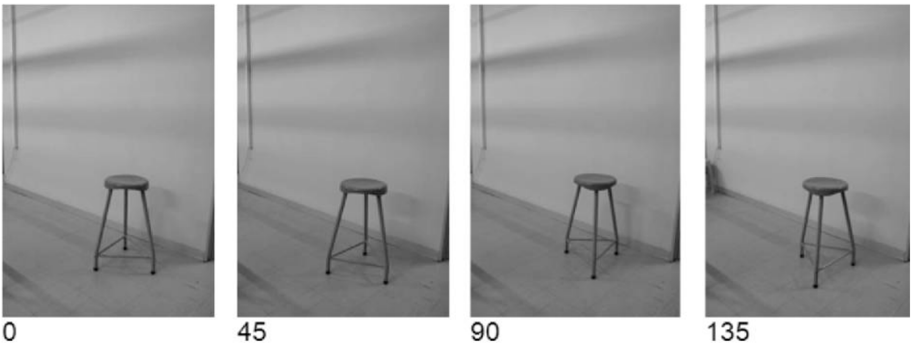


Fig. 2. Object A at rotations of 0, 45, 90, 135, 180, 225, 270 and 315 degrees.



180 225 270 315

Fig. 3. Object B at rotations of 0, 45, 90, 135, 180, 225, 270 and 315 degrees.

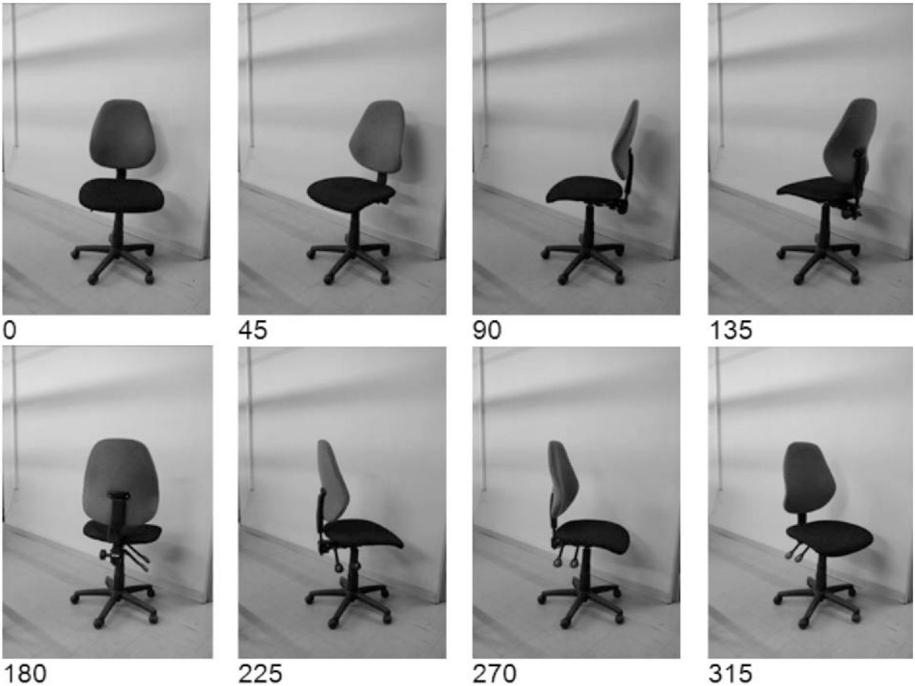


Fig. 4. Objects C at rotations of 0, 45, 90, 135, 180, 225, 270 and 315 degrees.

Preliminary experiments show that better results are achieved if the object is extracted form the background. It's noted that the background should be white as most pictures on the web has a white rather than a black background. A black background is normally generated with masking dew to the result of an mathematical image threshold function.

Custom extraction software is written and developed in a LabView environment. A threshold is applied to the image generating a binary mask. This binary threshold image mask is applied to the colour image. The resultant black background is replaced with a white background.

A

Fig. 5. Object A with and without the threshold function

B

Fig. 6. Object B with and without the threshold function

C

Fig. 7. Object C with and without the threshold function

The background replacement is done on each image before it is entered as a search.

4. IMAGE SEARCH FUNCTION AND RESULT VOICE COMMUNICATOR

For evaluation the experiment included:

- An image search of the object before processing without a descriptive.
- An image search of the object before processing with "office" as descriptive.
- An image search of the object after background replacement without a descriptive.
- An image search of the object after background replacement with "office" as descriptive.

The operation is as follows:

- The system opens the sample in Google's™ search bar and click search.
- The system would select the "Visually similar" option.
- The system saves the web page.
- The system would then read though the names (link names) of the image names and look for a common denominator.
- This would be the object description. A different direction might be to search for a common denominator in all the websites as a link from here, but positive results was acquired by simply using the image link names.

5. ARTIFICIAL COMPETENTLY CREATIVITY RESULT SEARCH FUNCTION AND VOICE COMMUNICATOR

As described above the system will read the information and interpret the web search. The common denominator is read aloud with a text to voice code.

The system would use the common denominator as a descriptive on http://dictionary.reference.com/. This webpage is saved. The system would read through it and the noun descriptive is communicated to the user with text to voice code.

These functions fulfill the definition of computational creativity as the ability of a program or computer to show human-level creativity and interaction.

6. RESULTS

Figure 1 shows the three samples for evaluation. Figures 2 to 4 show how these objects where rotated at 45 degree angles, generating more samples.

Object extraction, as shown in Figures 5 to 7, was also done at each angle.

Object recognition success is measured when the system was able to identify the object as a chair or synonym of a chair. The results are calculated as a percentage with the total out of 8 (the success rate of the 8 rotation samples).

Object description result or dictionary search result, success is measured if the definition of the object recognition was correctly defined, even if the identification of the object was incorrect. This would indicate the ability of the system as a pure descriptor. The results are calculated as a percentage with the total out of 8 (the success rate of the 8 rotation samples)

Table 1: Object A without the threshold function and without descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
A with no	0	in	0	1
threshold	45	s	0	1
and	90	view	0	1
No room	135	imagine	0	1
description	180	es	0	1
	225	visites	0	1
	270	imagine	0	1
	315	imagine	0	1
		Total %	**0**	**100**

Table 1 shows the results of the images of object A at 45 degree rotation intervals without the threshold function, i.e without replacing background with white. The image is added to Google™ search as is, without any descriptor. The image search result column show the word that the system thought was "seen" (common denominator). This is also the word the system would speak aloud. Note the there was a 0 % success rate for object recognition as none of the image search result was a chair or synonym of a chair.

The next part of the system would look for the meaning or definition of the object on the web. This had a 100% success rate if one only evaluates this function as a descriptor of the word as it results from the first part.

For example the result “view” (from the 90 degree angle answer) was searched by the system and the result was displayed and spoke out aloud as:

1. an instance of seeing or beholding; visual inspection.
2. sight; vision.
3. range of sight or vision: Several running deer came into the view of the hunters.
4. a sight or prospect of a landscape, the sea, etc.: His apartment affords a view of the park.

5. a picture or photograph of something: The postcard bears a view of Vesuvius.

This is a correct interpretation of the word; view.

Table 2: Object A without the threshold function and with a descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
B with no	0	wp	o	1
threshold	45	wp	o	1
but a	90	view	o	1
room	135	image	o	1
description	180	wp	o	1
(office)	225	view	o	1
	270	view	o	1
	315	from	o	1
		Total %	**0**	**100**

Table 2 show the results of the images of object A at 45 degree rotation intervals without the threshold function. The background was not replaced with white. This image is added to Google™ search with a descriptor of "office" (where the object was "seen"). As before the image search result column shows the word that the system thought it has "seen". This is also the word the system would speak out aloud. Note there was a 0% success rate for object recognition. As before the next part of the system would search for the meaning or definition of the object on the web. This had a 100% success rate as defined before.

Table 3: Object A with the threshold function and without descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
A with a	0	tech	0	1
threshold	45	cd	0	1
but no	90	large	0	1
room	135	tiare	0	1
description	180	shop	0	1
(office)	225	shop	0	1
	270	shop	0	1
	315	tiare	0	1
		Total %	**0**	**100**

Table 3 shows the results of the images of object A at 45 degree rotation intervals with the threshold function. The background was replaced with white. This image is added to Google™ search as is, without a descriptor. As before the image search result column shows the word that the system thought it has "seen". This is also the word the system would speak out aloud. Note there was a 0% success rate for object recognition. As before the next part of the system would look for the meaning or definition of the object on the web and this had a 100% success rate as defined before.

Table 4: Object A with the threshold function and with a descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
A with a	0	chair	1	1
threshold	45	chair	1	1
and	90	chair	1	1
room	135	white	0	1
description	180	chair	1	1
(office)	225	chair	1	1
	270	chair	1	1
	315	armchairs	1	1
		Total %	**87.5**	**100**

Table 4 shows the results of the images of object A at 45 degree rotation intervals with the threshold function. The background was replaced with white. This image is added to Google™ search with a descriptor of "office" (where the object was "seen"). As before the image search result column shows the word that the system thought it has "seen". This is also the word the system would speak out aloud. Note that with this combination there was an 87.5 % success rate for object recognition. Note that only the recognition at 135 degrees was wrong. A 100% recognition can be achieved with a slight rotation or variant of threshold function.

As before the next part of the system would look for the meaning or definition of the object on the web and this had a 100% success rate. For the result chair the following was found, displayed and spoken out aloud:

1. a seat, especially for one person, usually having four legs for support and a rest for the back and often having rests for the arms.
2. something that serves as a chair or supports like a chair: The two men clasped hands to make a chair for their injured companion.
3. a seat of office or authority.
4. a position of authority, as of a judge, professor, etc.
5. the person occupying a seat of office, especially the chairperson of a meeting: The speaker addressed the chair.

As for armchair the systems found, displayed and spoke out aloud:

1. a chair with sidepieces or arms to support a person's forearms or elbows.

This system was able to identify object A with an accuracy of 87.5 % when the background was replaced with white and a descriptor of where the object was "seen", in this case an office was added.

Remember it was the first time this object was "seen" by the system and no prior learning was done by this system itself. It took the unknown object, searched it on Google™ and was able to get an identification and description in a short time.

The same test was done on object B and C. The results are as follow.

Table 5: Object B without the threshold function and without descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
B with no	0	https:	0	1
threshold	45	s	0	1
and	90	https:	0	1
No room	135	gallery	0	1
description	180	https:	0	1
	225	https:	0	1
	270	perpignan	0	1
	315	thumbs.ebaystatic.com	0	0
		Total %	**0**	**87.5**

Table 6: Object B without the threshold function and with a descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
B with no	0	ui	0	1
threshold	45	home	0	1
but a	90	ui	0	1
room	135	partition	0	1
description	180	unit	0	1
(office)	225	ui	0	1
	270	home	0	1
	315	home	0	1
		Total %	**0**	**100**

Table 7: Object B with the threshold function and without descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
B with a	0	products	0	1
threshold	45	www.homedepot.com	0	0
but no	90	a	0	1
room	135	big	0	1
description	180	x	0	1
(office)	225	x	0	1
	270	shop.bals.co.jp	0	0
	315	shop.bals.co.jp	0	0
		Total %	**0**	**62.5**

Table 8: Object B with the threshold function and with a descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
B with a	0	chair	1	1
threshold	45	chair	1	1
and	90	chair	1	1
room	135	chair	1	1
description	180	chair	1	1
(office)	225	chair	1	1
	270	chair	1	1
	315	chair	1	1
		Total %	**100**	**100**

Table 9: Object C without the threshold function and without descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
C with no	0	wroclawczyk	0	0
threshold	45	of	0	1
and	90	hotel	0	1
No room	135	hotel	0	1
description	180	octmonth	0	0
	225	s	0	1
	270	file.mafengwo.net	0	0
	315	products	0	1
		Total %	**0**	**62.5**

Table 10: Object B without the threshold function and with a descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
C with no	0	ui	0	1
threshold	45	chair	1	1
but a	90	space	0	1
room	135	space	0	1
description	180	space	0	1
(office)	225	ui	0	1
	270	chairs	1	1
	315	wp	0	1
		Total %	**25**	**100**

Table 11: Object C with the threshold function and without descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
C with a	0	i	0	1
threshold	45	shop	0	1
but no	90	tag	0	1
room	135	pumps	0	1
description	180	vyr	0	0
(office)	225	mcquade.com	0	0
	270	kch	0	0
	315	i	0	1
		Total %	**0**	**62.5**

Table 12: Object C with the threshold function and with a descriptor

Object	Rotation	Image search result	Object recognition Success (true =1, false =0)	Dictionary search result Success (true =1, false =0)
C with a	0	chair	1	1
threshold	45	armchairs	1	1
and	90	chair	1	1
room	135	chair	1	1
description	180	chair	1	1
(office)	225	chair	1	1
	270	chair	1	1
	315	chair	1	1
		Total %	**100**	**100**

Tables 5 to 12 show that the system was able to identify object B and C with an accuracy of 100%, only when the background was replaced with white and a descriptor was added of where the object was "seen".

Note as before that it was the first time this object was "seen” by the system and no prior learning was done by this system itself. It took the unknown object, searched it on Google™ and was able to get an identification and description in a relatively short time.

Note that the dictionary search was only valid if a valid word was a result at image search result column. For instance wroclawczyk (as found at C with no threshold and no room description at 0 degrees) is just gibberish. The system stated that no result was found. This can be seen as a positive answer as the system came to the result that this is not a word.

7. CONCLUSION

The positive results in this experiment indicate that with novel interpretation software it is possible to use Google™ as base for an artificial intelligent, computational creative system. This is achieved by programming an expert system as an interpreter of online content. Philosophically this is identical to its human counterpart. If we as humans do not know an answer we would “Google™” it. As stated computational creativity is the ability of a program or computer to show human-level creativity and interaction. Optimal results of computational creativity can only be achieved if such an intelligent system is trained on human data for optimal human like mimicking. Information on the web is “human upload”, giving the system human intelligence. This human uploaded data became apparent in that the height from where the uploaded pictures were taken is on human eye level. This must be taken into consideration when mounting the cameras.

The system has its shortcomings and the best result can be achieved when a keyword of where an object was “seen” is added. It should however be possible to identify a room, by doing an image search of it in the same way as the object search is done.
These results indicate that it is possible to build truly incredible human trained creative computation artificial intelligent expert systems by using web as intelligence base concentrating on interpretation capability.

8. FUTURE RESEARCH

It should be possible to identify any object of which there is a picture on the web. If such a picture does not exist one can upload it do the web for future reference and by doing so inadvertently building a world wide database of these unknowns. Preliminary test showed promising results.

One need not be limited to pictures and should also be able to interpret a video clip. For example one could ask a system to make coffee. It should be able to identify the objects on the table like the mug, kettle, sugar, spoon and coffee. Using a web search the system should be able to get instructions and a video how to do it.

The positive creative computation results showed that it is possible for the system to give a description of objects. Web searches also aid creative computation as a chatterbot would be able to ask you for instance where you work, search this and generate a conversation. It might be able to identify you on your Facebook page and generate conversation on this.

9. REFERENCES

Duch, W. 2006, Computational Creativity IEEE International Joint Conference on Neural Networks Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada July 16-21, 2006. p. 435

http://www.analyticspress.com/datacenters.html August 1, 2011 downloaded on 2012/07/19, p.3,10,13

Koomey, J, G. 2011, GROWTH IN DATA CENTER ELECTRICITY USE 2005 TO 2010 A report by Analytics Press, completed at the request of The New York Times

Mack, E. 2011. Google™ Wins the War Against Bing Images

Miller, G. Fels, S. Oldrige, S., 2011, A Conceptual Structure for Computer Vision Canadian Conference on Computer and Robot Vision 2011 IEEE. p.1.

Ooh! Ahh! Google™ Images presents a nicer way to surf the visual web, http://googleblog.blogspot.com/2010/07/ooh-ahh-google-images-presents-nicer.html Google official blog, Downloaded on 2012/07/19

Pereira, F, C., 2007, Creativity and Artificial Intelligence: A Conceptual Blending Approach, Walter de Gruyter GmbH & Co. Germany. P.205-p.207

Slack J, M. 1991, Hybrid Encoding: Constraints on addressing structure, Trends in Artificcial intelligence: 2nd congress of the Italian association for Artificial intelligence, Ai*IA Palermo, Italy, October 1991 Proceedings, Springer- Verlag Berlin Heidleberg. p.117

Szeliski, R., 2011, Computer Vision: Algorithms and Applications.Springer Verlag London Limited. p.5