Automated Collage Generation – With More Intent

Michael Cook and Simon Colton

Computational Creativity Group
Department of Computing, Imperial College, London, UK.

ccg.doc.ic.ac.uk

Abstract

The majority of software has no meta-level perception of what it is doing, or what it intends to achieve. Without such higher cognitive functions, we might be disinclined to bestow creativity onto such software. We generalise previous work on collage generation, which attempted to blur the line between the intentionality of the programmer and that of the software in the visual arts. Firstly, we embed the collage generation process into a *computational creativity collective*, which contains processes and mashups of processes, designed so that the output of one generative system becomes the input of another. Secondly, we analyse the previous approach to collage generation to determine where intentionality arose, leading to experimentation where we test whether augmented keyword searches can enable the software to exert more intentional control.

Introduction

We are building The Painting Fool software (www.thepaintingfool.com) to one day be taken seriously as a creative artist in its own right. To guide this process, we have engaged in much informal discussion with artists, art students and teachers. One clear opinion almost universally expressed has been that art generation software such as Photoshop is not seen as creative, because it exhibits no intentionality. That is, it does not conceive the artwork it wishes to produce, and hence does not drive the process through aesthetic or production decisions. Rather, the person using Photoshop is seen as providing all the intentionality, and hence is the sole creative entity, with the software acting as a mere tool.

To address this lack of intentionality in The Painting Fool, as described in (Krzeczkowska et al. 2010), we built a collage-generation module able to construct artworks to illustrate newspaper articles. The system worked via internet retrieval of a news article, text extraction of important keywords, internet image retrieval using the keywords, and non-photorealistic rendering of the images. We removed human decision making, so that we had no control over (a) what news story would be chosen for a collage (b) what keywords would be extracted (c) what images would be retrieved or (d) how the collage would be rendered. We used the system to raise the issue of whether it was appropriate to say the software exhibited intentionality in producing the collages.

We describe here two extensions to this previous work. In the next section, we show how the collage generator can be seen as a *mashup* of generative processes, and we give details of a computational creativity collective designed with the hope that by allowing the output of such processes to become the input of others, more culturally interesting artefacts will be produced. Following this, we return to the question of intentionality, and analyse the results of the existing collage generation system. This highlights the roles that five parties have in adding intentionality to the generation process, and suggests simple improvements which could wrestle back some control for the software. We present the results from some preliminary experimentation with augmented keyword image retrieval, which leads to further discussion about intentionality in software. We conclude by describing avenues for future work.

A Computational Creativity Collective

The software discussed in (Krzeczkowska et al. 2010) is best described in modern computing parlance as a *mashup*. As described in (Abiteboul, Greenshpan, and Milo 2008), in general, mashups take data – usually downloaded from internet sites – and pass them through various linked processes in order to turn raw data into more consumable (amusing/entertaining/informative/interactive) presentations. In contrast, in general, the kinds of creative systems that produce artefacts like musical and literary compositions, works of visual art, scientific hypotheses, etc., are run as standalone, offline, systems. We hypothesise that the value of the artefacts produced by creative systems, and hence the creativity attributed to them, would be increased if (a) they could place their work into current cultural contexts and (b) they could work in concert, whereby the output from one becomes the input to another, so that there is an increase in the sophistication of the overall system.

To test this hypothesis in the long term, we have compiled a collective of processes and mashups employing these processes, which is available at (www.doc.ic.ac.uk/ccg/collective), along with a simple Java API which guides the construction of additional material for the collective. Currently, the collective contains 119 processes mashed-up into 57 mashups. The processes follow an interface which means that the output from any process can become the input to any other (whether or not

the receiving process can usefully work with the type of input it is given). In general, the current processes perform simple text or graphics routines, or they employ a web 2.0 API in order to retrieve data from the internet. Collage generation is represented with a mashup that contains processes which link with the Guardian newspaper API and the Google image retrieval API, in addition to processes which perform text extraction and graphics routines. The other mashups perform similarly. For instance, one mashup uses image retrieval, graphics, and natural language generation techniques to parody business speak. Another mashup uses processes based on the LastFM and RunKeeper APIs, so that a mosaic of album covers can be constructed depending on what music people run the fastest to. Such mosaics are not necessarily the most culturally important artefacts produced by creative systems. However, the ease by which such mashups can be written highlights the potential for creativity when story generation systems feed into poetry generators, the output of which inspire paintings that influence musical compositions, and so on, with each process benefitting from internet downloads to add timely cultural relevance. We discuss our aims and future directions for the collective in the final section below.

Intentionality Analysis and Experimentation

One of the most impressive collages presented in (Krzeczkowska et al. 2010) is given in figure 1. This was produced in response to a news article which covered the war in Afghanistan, published in 2009 in the Guardian newspaper. While not rendered to a particularly high aesthetic standard, the contents are striking, as it contains a bomber plane, an explosion, a family with a baby, a girl in ethnic headgear, and – most poignantly of all – a field of war graves. By mixing images of death and destruction with those of children, it is fair to say that these constituents have helped to produce a collage with a moderate but definite negative bias, which reflected the tone of the original news article. We previously thought that the intentionality driving the collage generation therefore came from four parties: (i) the programmer, by enabling the software to access the left-leaning, largely anti-war Guardian newspaper (ii) the software, through its processing, the most intelligent aspect of which was to extract keywords using an algorithm based on (Brin and Page 1998), (iii) the writer of the original article, through the expression of his/her opinions in print, and (iv) individual audience members who have their own opinions forming a context within which the collages are judged.

However, upon further inspection, we can add a fifth party to this ensemble. To see this, note first that the keywords extracted from the article were:

afghanistan, brown, forces, troops, nato,
british, speech, country, more and afghan

We further note that none of these words have a particular negative emotive bias. We therefore must reduce the intentional role of the article writer in the construction of the collage, as it would be possible to write a very upbeat appraisal of the war which contained the same keywords. The images for the collage were retrieved from Flickr using the
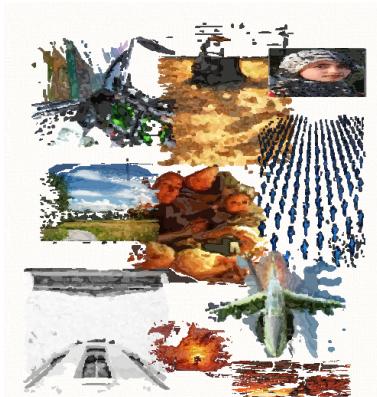
Figure 1: Collage illustrating the war in Afghanistan

keywords above. Hence, it is clear that the negative connotations in the collage derive from the way in which Flickr users have tagged the images they have uploaded, presumably by tagging images of cemeteries and explosions with words like 'afghanistan', 'nato' and 'troops'. Hence, we must attribute some of the intentionality behind the collage construction to the *crowd*.

It is interesting that intent can be dissipated amongst five parties, one of which is a faceless crowd. However, the software is currently the junior partner in this arrangement, and our original motivation was for it to become the major driving force, so that it might be perceived as being more creative. We are currently working on an opinion-forming module for The Painting Fool which will enable it to start with some text, such as a news article or a set of search terms, then download and analyse multiple texts on related subjects, in order to form either a positive or negative opinion about the original text. Once it has formed such an opinion, this will be used to bias the search for images in collage generation, i.e., if the opinion is positive, The Painting Fool will attempt to retrieve appropriately up-beat images, and vice-versa if not.

A straightforward method for attempting to influence the emotional nature of retrieved images is to augment the keywords with appropriately positive/negative adjectives. To test the effectiveness of this method, we performed some preliminary experiments. We used augmented keyword searches with Flickr to retrieve images, and then asked people to tag the images with respect to the broad emotion being portrayed. In particular, in the first experiment, 17 participants were each asked to tag 21 images, 7 of which were retrieved from Flickr with the keyword *soldier*, 7 of which were retrieved with the keywords: *soldier* and *sad* and 7 of which were retrieved with the keywords: *soldier* and *happy*. For each set of keywords, 80 images were retrieved and cached, and these were selected from randomly during the experiment. Subjects were asked to tag each image as either 'happy', 'sad' or 'unsure' (if they were not prepared to assign an emotion to the image). The same experiment was repeated, but with the keyword *soldier* replaced by *dog*, and repeated again with the keyword *baby*.

A sample of the images from the dog experiment is given in figure 3, and the distribution of tags for the sets of images
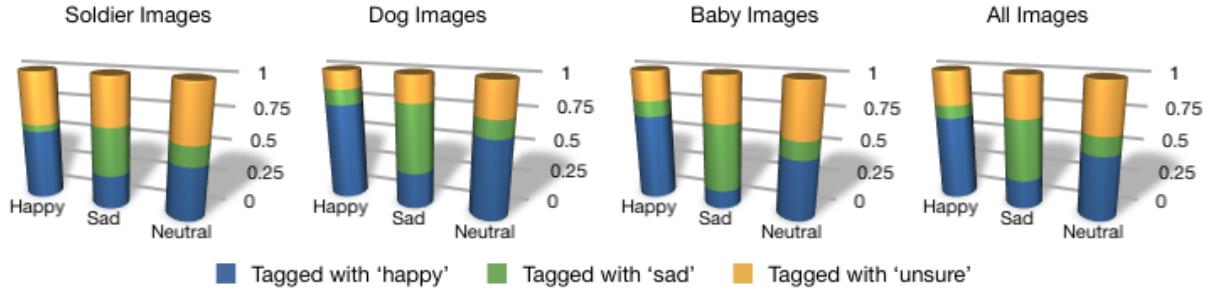
Figure 2: Results from the image tagging experiments. *Neutral* refers to the images retrieved via unaugmented search.
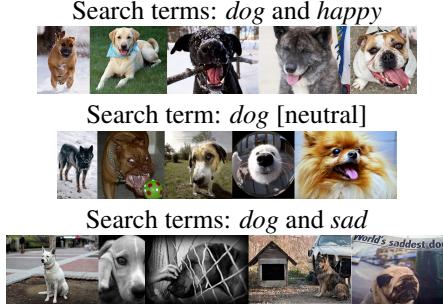
Figure 3: Sample images from the dog image tagging exp.

in the three experiments is given in figure 2. We see that the participants' reactions to images were image-type dependent. In particular, people were less sure about the emotions being portrayed in the soldier images than in the other two types of image. Also, the results show a trend for neutral images to be tagged with 'happy' as often as 'unsure' and four times as often as 'sad'. It's likely that this is because people tend to upload pictures of happy babies/dogs/soldiers to Flickr, hence an unaugmented search will result in largely positive images being retrieved. Importantly, it is clear that people were more likely to tag images retrieved with the keyword *happy* as 'happy' and with the keyword *sad* as 'sad'. Of the 119 *happy* images, 65% were tagged with 'happy' and only 10% with 'sad'. Similarly, of the 119 *sad* images, 47% were tagged with 'sad' and only 22% with 'happy'.

Conclusions and Future Work

The Painting Fool already has access to the methods and mashups in the computational creativity collective, including the collage generator mashup, and we plan to explore the creative potential of this. In particular, given that any process in the collective can provide input to any other process, we plan to automate the construction of chains of processes to see what the resulting mashups produce. We also plan to explore more sophisticated ways of using the processes, for instance using global workspace architectures, as explored in (Charnley 2009). Our hope is that the collective will become a resource to which computational creativity researchers can contribute creative systems, in addition to experimenting with those of others. Moreover, we envisage the collective becoming a creative entity in its own right, producing artefacts of real cultural value.

The results from the preliminary image tagging experiments described above were encouraging. Hence, in future versions of the software, when The Painting Fool forms an opinion about a subject, then attempts to express that opinion via collages built using augmented keyword searches, we can be fairly confident that the attempt will be successful. However we will need to undertake further experimentation to determine under which conditions this will be the case. The opinion forming process will rely on sentiment analysis routines, such as those described in (Pang, Lee, and Vaithyanathan 2002), but The Painting Fool will be trained to subvert popular sentiment on particular topics, or attempt to portray both sides of an argument, if either of these methods might produce artworks with greater impact. Currently, the collages it produces are too literal. Hence, we plan to implement some obfuscation techniques, which will be used to produce pieces which require a level of interpretation by audiences. By implementing such higher level abilities, we hope to endow The Painting Fool with behaviours that will one day be regarded as creative.

References

Abiteboul, S.; Greenshpan, O.; and Milo, T. 2008. Modeling the mashup space. In *Proceedings of tht 10th ACM International Workshop on Web Information and Data Management*.

Brin, S., and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. *Comp. Networks & ISDN Syst.* 30.

Charnley, J. 2009. *A Global Workspace Framework for Combined Reasoning*. Ph.D. Dissertation, Department of Computing, Imperial College, London.

Krzeczkowska, A.; El-Hage, J.; Colton, S.; and Clark, S. 2010. Automated collage generation - with intent. In *Proceedings of the 1st International Conference on Computational Creativity*.

Pang, B.; Lee, L.; and Vaithyanathan, S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proc. of the Conf. on Empirical Methods in Natural Language Processing*.