

PicWords: Creating Pictures by its keywords

Ankur Singh¹, Shruti Garg²

¹ M.Tech, Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India.

² Assistant Professor, Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India.

Abstract - PicWords can be considered as a kind of Non-Photorealistic Rendering (NPR), more specifically belonging to a subclass, called NPR Packing, which focuses on re-creating an image by arranging a collection of small pictorial elements. The pictorial elements can obviously be small images, tiles, or text. The most fascinating part of PicWords is that it attracts the human visual system using two modalities, i.e. picture and text (keywords). Once recognizing the picture, on close inspection, the general elements, i.e. the keywords can be noticed. The entire generation process of PicWords consists of four modules: a picture module, a keywords module, a cross modality picture and keywords module and a final post processing module. The output of picture module is ranked patches, which act as containers for the output of the keywords module, i.e. the keywords. In the picture and keywords module, a direct correspondence is made between these patches and keywords based on their ranks, by warping individual keywords along the contour of the image patches. The post processing module refines the output of picture and keywords module for better visual perception. After the post processing module, the final output, the PicWords, is obtained. Experiments were performed on various images and it has been found that the techniques used in this work provide better results than in previous works.

Key Words: PicWords, Non-Photorealistic packing, calligram, keyword, picture.

1. INTRODUCTION

WHEN a person sees an attractive picture, say, a photo of some celebrity like Aishwarya Rai, they may want to know more details about the celebrity. Since the picture is no means of telling any such details like their date of birth, nationality, achievements, etc., they may have to search such details on some other source of information, like a website, or the person's blog. If a person otherwise reads a piece of information about the concerned personnel on some source, he may revert to some other source to figure out how the celebrity looks like in person. This kind of dilemma is commonly seen with people in day to day life.

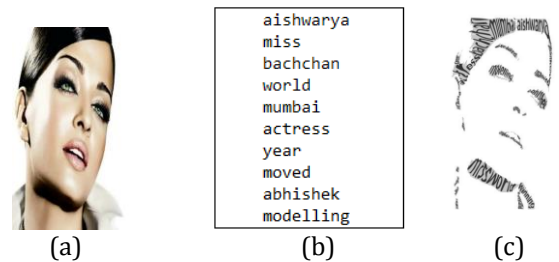


Fig -1: Output of PicWords (a) Original image (b) keywords (c) Final output

Now suppose a system is designed to produce an image, which contains said celebrity's face, and at the same time, contains a brief summarization of their credentials. One such system, PicWords is shown in Fig-1. It comprises of Aishwarya Rai's face, as well as contains certain keywords which introduce her in a brief manner. In PicWords, two modalities, i.e. the picture and the keywords, are fused together to represent the object of interest in a better way. The two modalities are complimentary to each other, and contribute to a better understanding of the image in an artistic way. The image helps in better recognition of the object of interest and the keywords makes the art-form rich in information. As mentioned earlier, PicWords is an example of Non-photorealistic rendering, more precisely belonging to the category of NPR Packing, which focuses on re-creating an image by arranging a collection of small pictorial elements. The pictorial elements can obviously be small images, tiles, or text. The most fascinating part of PicWords is that it attracts the human visual system using two modalities, i.e. picture and text (keywords). Once recognizing the picture, on close inspection, the general elements, i.e. the keywords can be noticed.

The entire generation process of PicWords consists of a picture module, a keywords module, a cross modality picture and keywords module and a final post processing module. The output of picture module is ranked patches, which act as containers for the output of the keywords module, i.e. the keywords. In the picture and keywords module, a direct correspondence is made between these patches and keywords based on their ranks, by warping individual keywords along the contour of the image patches. The post processing module refines the output of picture and keywords module for better visual perception.

In section 2, the related work has been. Then, the whole system framework and methodology is illustrated in section 3. Next, in section 4, the current state and progress of the system is presented. Results and conclusion of the thesis are shown in section 5 and the future aspects of the current work are shown later in section 6.

2. RELATED WORK

Research in the field of Computer Graphics has traditionally been characterized by attempts to achieve photorealism; modeling physical phenomena such as light reflection and refraction to produce scenes lit, ostensibly, in a natural manner. Over the past decade the development of novel rendering styles outside the bounds of photorealism has gathered momentum: the so called non-photorealistic rendering or NPR. NPR has been successfully applied by scientific visualization researchers to improve the clarity and quantity of information conveyed by an image; exploded diagrams in maintenance manuals and false color satellite imagery are two common examples of such visualizations. Several NPR techniques have been designed to emulate a broad range of artistic media from pastel to paint, and to mimic many artistic techniques such as hatching and shading. Non-photorealistic rendering, recently is subject to increased efforts and devotion, especially its subclass named NPR packing. It is based on arranging small tiles, text or other pictorial elements to create artistic art forms so as to enhance multimedia presentations. NPR packing can be further sub classified into mosaicking and calligrams. Mosaicking, which is also popular by the name tiling or stippling aims to recreate the image using a medium, which is usually tiles or elementary pictorial element and packing image regions with those miniscule atomic rendering elements.

Besides mosaicking, some artists have also succeeded in combining text words and images, and hence developed the second art style called calligram. Calligram is the recreation of a target image by arranging an array of small text/words, each chosen specifically to fit a particular block/container of the target image.

2.1 Mosaicking: Years of research has led to some amazing work in the field of NPR packing, especially mosaicking. Mosaics are images made by cementing together small colored tiles. The creation of digital mosaics of artistic quality is one of the challenges of the Computer Graphics and is one of the most recent research direction in the field of Non Photorealistic Rendering. Digital mosaics are illustrations composed by a collection of small images called "tile". The tiles "tessellate" a source picture in order to reproduce it in a "mosaic-like" style. Starting from the same source image it is possible to create different kind of digital mosaics depending on the choice of the tile dataset and the imposed constraints for positioning, deformations, rotations, etc. In particular it is possible to put the mosaic building from a source raster

image in terms of a mathematical optimization problem as follows:

"Given a rectangular region I^2 in the plane R^2 , a tile dataset and a set of constraints, find N sites $P_i(x_i, y_i)$ in I^2 and place N tiles, one at each P_i , such that all tiles are disjoint, the area they cover is maximized and the constraints are verified as much as possible."

Based on this definition, Digital mosaics can be further classified into crystallization mosaics, ancient mosaics, photo mosaics and puzzle image mosaics.

(a) Crystallization Mosaics: Paul Haeberli [5] used Voronoi diagrams, placing the sites at random and filling each region with a color sampled from the image. This approach tessellates the image and tile shapes are variable and do not attempt to follow edge features; the result is a pattern of color having a cellular-like look. Dobashi et al. [6] reprised the Haeberli's idea obtaining better results due to the fact that they address the problem of keeping into account the edges of the original image. However the tile shapes suffer from the extreme variability of the Haeberli's technique. Faustino and Figueiredo [7] presented a technique similar to the Dobashi et al. approach, in which the main difference is that the sizes of tiles vary along the image: they are small near image details and large otherwise.

(b) Ancient Mosaics: Hausner et al. [10] were the first to having simulated the appearance of Roman mosaic. Their work which has addressed shape packing using irregular tile shapes, through an efficient energy minimization scheme. Elber and Wolberg [11] came up with a very advanced technique to render traditional mosaics. This technique is based on offset curves that get trimmed-off the self intersecting segments with the guidance of Voronoi diagrams. More approaches for the creation of ancient mosaics are presented by Battiato et al. [9] and Di Blasi and Gallo [8]. These approaches are based on directional guidelines, distance transform, mathematical tools and century proved ideas from mosaicists and leads to impressive results. Recently a novel technique for ancient mosaics generation has been presented in [12]. The authors present an approach for stroke-based rendering that exploits multi-agent systems; they call the agents RenderBots. RenderBots are individual agents each of which in general represents one stroke. They form a multiagent system and undergo a simulation to distribute themselves in the environment. The environment consists of a source image and possibly additional G-buffers. The final image is created when the simulation is finished by having each RenderBot execute its painting function. In another approach, Lloyd's method was applied to Voronoi diagrams effectively utilizing the Manhattan metric, hence giving an arrangement of oriented regular rectangular tiles.

(c) Photo mosaics: Photomosaic is one of the most interesting technique (and one of the most cloned algorithm) in the field of digital mosaic. Photomosaic

transforms an input image into a rectangular grid of thumbnail images. In this approach the algorithm usually searches in a large database of images for one that approximates a block of pixels in the main image. The resulting effect is very impressive. In 1973 Leon Harmon [13] presented a work including several "block portraits". Harmon used these "pixelated" portraits to study human perception and the automatic pattern recognition issues. For example he used them as a demonstration of the minimal condition to recognize a face. In 1976 Salvador Dali [14] completed the painting titled "Gala contemplating the Mediterranean Sea, which at 30 meters becomes the portrait of Abraham Lincoln (Homage to Rothko)". Note that Lincoln's face is made up by pictures with full tonal ranges; perhaps the earliest example of this technique. The nude taking up several tiles is Dali's wife Gala and one of the tiles is Harmon's gray scale image of Lincoln; so not only Dali did an appropriate Harmon's Lincoln portrait into the overall composition, but he also reincorporated a smaller gray scale version into a single tile. If not the first, this is certainly the most well known image made from other images. In the 1970's the American artist Chuck Close [15] began producing precisely gridded paintings. Some of these had a quality that reminds Impressionists, while others seem to be computer generated or computer influenced. The earliest example of a photographic computer mosaic is the image created by Dave McKean. Computer replicated images made from tiled digital photographic pictures is recent because it requires large computational resources. Robert Silvers [16] began working on the first photomosaic while he was a graduate student at the MIT Media Lab. Each tile in his images represents much more than a single value. Smaller pictures match the overall image in tone, texture, shape and color. Silvers was commissioned by the US Library of Congress to create the portrait of Lincoln using archived photos of the American Civil War. Today, Runaway Technology [17] (Silvers' company) produces photomosaic images as logos and illustrations for individuals, corporations and publications. William Hunt [18], a computer programmer, created the image. He used three different size tiles to change the look of the grid. Scott Blake's image of Abraham Lincoln is rendered by using 42 portraits of all US Presidents [20]. He arranged the presidents according to their gray scale density. He offset the tiles on a beehive grid pattern to produce a hypnotic effect and used the oval portraits to fill the space in a more efficient manner.

(d) Puzzle mosaics : Puzzle Image Mosaic is inspired by Giuseppe Arcimboldo [23], a Renaissance Italian painter inventor of a form of painting called the composite head where faces are painted not in flesh, but with rendered clumps of vegetables and other materials slightly deformed to better match the human features. Kim and Pellacini [25] presented Jigsaw image mosaics (JIM) approach, using an active contour based optimization scheme to minimize the energy function that traded off

among various measures of the packing's quality. Another approach for the creation of the same kind of mosaics is presented in [24]; this approach is based again on the Antipole strategy and leads to impressive results in an acceptable computation time. The technique reformulates the problem as a search problem in a large database of small images and takes into account some important features of the image to speed up the search process. Orchard and Kaplan [26] described a fast technique for mosaicking images with irregular tiles, also capable of cropping partial regions from the image database to use as tiles. Although the previous approaches have been great success in achieving desired mosaics, it is very difficult for the viewer to read and comprehend the content related to the target image. On the contrary, PicWords is rather more informative than mosaicking, by effectively inserting the keywords as the packing elements into the target image, therefore allowing the viewers to sense the visual how abouts of the image, as well as have a brief view of the key object inside the image.

2.2 Calligrams: Another class of NPR packing, calligrams has been studied in a number of contexts. A calligram is an arrangement of words or letters, designed to create a visually perceivable image. Calligrams have a rich tradition and wide variety of styles depending on the imagination of the artist. ASCII art is one such example, which is a technique of recreating pictures with ASCII text characters. In ASCII art, textual and numeric characters can be used to recreate an image; that is, the single characters which are not supposedly meant to have any meaning, can be packed together to be perceived as components to form a whole. Xu [27], in his work showed the generation of ASCII art which was purely structure based by analyzing the contour structure of the image. Structure based ASCII art captures the major structure of the image content. Nacenta [28] developed a technique called FatFonts based on Arabic numerals. This enables accurate reading of the numerical data while preserving an overall visual context. The drawback shared by the above mentioned systems is that no relationship exists between the target image and its components (text or Arabic number). Xu and Kaplan [2] developed a system for packing letters into images, a specific case of irregular tiling. They decomposed the image into sub parts, called containers, and effectively warped text into those containers. But their method can only work well with letters, and thus cannot be considered to convey a lot of meaningful information. Later, Maharik and Sheffer [29] presented an algorithm for creating images using miniscule text, and called it digital micrography. These attractive text-art works effectively combine beautiful images with meaningful text. The drawback to their approach is that the text is miniscule and sometimes it is hard to read and comprehend them. Hu and Liu [1] proposed a novel text-art system, input a source picture and some keywords introducing the information about the

picture, and the output is the so-called PicWords in the form of the source picture composed of the introduction keywords. Different from traditional text-graphics which are created by highly skilled artists and involve a huge amount of tedious manual work, PicWords is an automatic non-photorealistic rendering (NPR) packing system. However, their work suffers computational delay due to the fact that segmentation of the image is time consuming. Also, since the system takes contour of various elements into account and replaces the original texture of a region with a keyword, it will probably fail on textureless images.

3. SYSTEM OVERVIEW

This section contains an overview of the PicWords system. The system is divided into four modules: picture, keywords, picture and keywords and post-processing. The first part is picture module. Given a source image, first it's silhouette image is generated. A silhouette image is the dark shape and outline of someone or something visible in restricted light against a brighter background. In the silhouette image, unnecessary background details are removed and only the important patches are kept. Then the silhouette image is segmented into several small patches with the Simple Linear Iterative clustering (SLIC) algorithm [34]. Since only the relevant foreground area is to be occupied, only patches covered by the silhouette image are kept. Once the patches have been generated, they are ranked according to their capacity to contain a keyword. Usually, central and long patches are supposed to rank higher. This is because while looking at an image, a viewer focuses more on central patches. So, longer and more relevant keywords should be kept in central and longer patches. In the keywords module, first relevant and required introduction from Internet (Wikipedia, Twitter, etc.) is extracted. After filtering stop words and adjusting their case, the keywords are assigned a weighted rank according to the keywords frequencies and their length. In the picture and keywords module, each word in the keywords list corresponds to its counterpart in the patch list. Higher weighted keywords should fit into more salient, larger patches because longer keywords in central area can better capture the viewer's attention. In the post-processing module, several kinds of strategies can collaboratively help generate more meaningful PicWords.

3.1 TECHNIQUE DETAILS OF THE PICWORDS SYSTEM

In this Section, the techniques used in picture module, keywords module, picture and keywords module and post-processing module have been discussed sequentially.

3.1.1 Picture Module: A flowchart showing the basic steps in picture module is shown in Fig- 2.

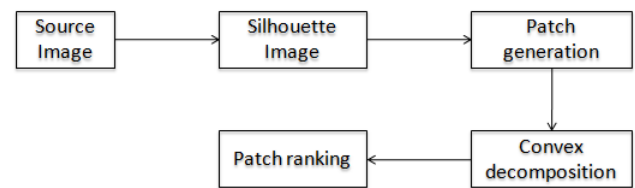


Fig- 2: Basic steps in picture module

- a) **Silhouette Image Generation:** First of all, the original image is split up into two parts, a yang part and a yin part. The Yang part corresponds to important image content, in which keywords are to be placed. On the other hand, the yin part is often unnecessary trivial background details which are not to be considered, and hence is filtered away. In previous works, mean shift technique [31] has been used to segment the image into small superpixels. Considering the noticeable amount of computational time required to perform mean shift, a rather less costly method is used in this work. The fuzzy c-means approach [35] is used to segment the image. Since the keywords are to be placed in the foreground i.e. the object of interest, the background is filtered off and thrown away. Background can be removed by many methods, like lazy snapping technique used in [36]. The foreground is also segmented simultaneously. Spatially nearby areas having similar pixel intensities, like the hair or the cheek are kept together in same cluster. Once this has been done, the segmented foreground is converted to grayscale image, which is then further thresholded into a binary image. This threshold can be decided by the user. Finally, a Gaussian filter is used to smoothen up the binary image, and to take care of the irregularities like holes in the image to generate the final silhouette image. The final silhouette image of an original image is shown in Fig- 3(b).

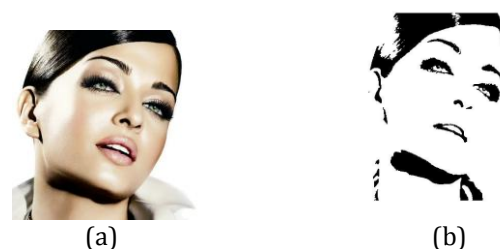


Fig- 3: Silhouette image generation using Fuzzy c-means segmentation (a) original image (b) silhouette image

- b) **Patch Generation:** Once the desired silhouette image has been generated, images need to be decomposed into containers in which keywords will be placed later. This should be done in such a way that the entire image is captured without compromising the boundary details of the image. For this, the image must be broken into smaller superpixels. A superpixel is a near- polygonal sub-part of a digital image. It is larger than a normal pixel, and is

rendered in the same color and brightness. In other words, a superpixel can be defined as a group of pixels which have similar characteristics. There are many approaches for generating superpixels, each with its own advantages and drawbacks, and each one of them is better suited for different types of problems. It is difficult to find out a single technique to work with all these problems, but following properties are generally desirable while creating superpixels:

- (a) Superpixels should adhere to the boundaries of the image. The boundary of a superpixel should not coincide or run over the image boundary.
- (b) Superpixels should be easy to generate, computationally fast, memory efficient and simple to use for a normal user.
- (c) Superpixels should not decrease the quality of an image.

Hu et al. [1] have used state of the art superpixel segmentation algorithm [32] to generate superpixels. Achanta et al. [34] has performed comparisons between five state of the art superpixel algorithms based on the above mentioned criteria of speed, computation time and adherence to image boundary. She has concluded that no particular method is satisfactory for all applications. To address the problem of computational cost, speed, and ease of use, Achanta et al. [34] came up with a new superpixel method, called Simple Linear Iterative Clustering (SLIC) algorithm, which is an extension of the already existing k-means clustering to generate superpixels. While being very simple, SLIC considers the problems of computational speed, cost and adherence to image boundaries, and out performs other state-of-the-art techniques in terms of simplicity, flexibility and compactness of superpixels. It takes the required number of superpixels as input and gives the image decomposed into superpixels. The silhouette image is fed to SLIC algorithm as an input, along with the desired number of superpixels 'k'. The algorithm iteratively generates an image composed of 'k' superpixels. The output is shown in Fig- 4.

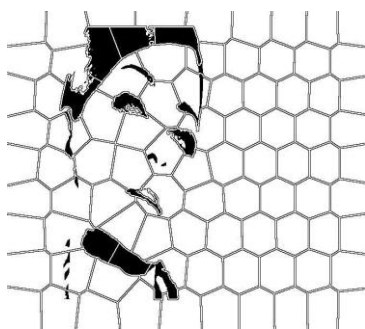


Fig- 4: Patch generation using SLIC superpixel segmentation with 100 superpixels.

c) Patch Ranking: Not all the generated patches are equally important. For example, larger patches near the center are more important. Patches are ranked based on two criteria: first, patch area S_i , which is the number of pixels that are contained within the patch region; second, patch location D_i , indicating the distance of the patch from the centre of the image and third, length of major axis of the patch, L_i . Based on these criteria, the weight of a patch is presented as:

$$W_i = W_1 S_i - W_2 D_i + W_3 L_i$$

Where W_i is the weight coefficient of each element i . This coefficient can be manually decided by the user and can vary for different images.

3.1.2 Keywords Module: There are mainly two parts in the keywords module. One is keywords collection and the other is keywords ranking. A flowchart of keywords module has been shown in Fig- 5(a).

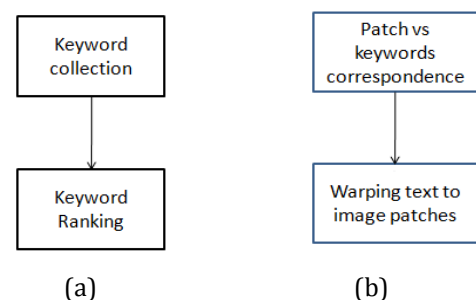


Fig- 5: Basic steps in (a) keywords module (b) picture and keywords module

a) Keywords Collection: The keywords are searched from Internet, using some reliable source such as Wikipedia, Twitter, Weibo, etc. and processed to remove peculiars and stop words. Also, text lowercasing or uppercasing can be done, depending upon the need of the user. Following keywords about the object of interest (Aishwarya Rai in this case) have been taken from IMDB. A paragraph containing the information of the object of interest is analyzed and words are stored in form of an array. Before storing the words in an array, the stop words are removed. Once the keywords have been stored, a count is made to find out the frequency of occurrences of all the words. Only the words with a certain minimum length are considered so as to remove all the adjectives. After that, the percentage of occurrence of each word relative to the total number of words in the paragraph is computed, and keywords are given rank according to the obtained frequencies. Keywords having more occurrences than the others are given relatively higher rank, and those keywords which are less frequent are given lesser rank. Once this has been done as shown in Fig- 6 and 7, the keywords crawled from various internet sources and processed by lowercasing and stop words removal are ready to be fed into the patches generated in the picture module.

Aishwarya Rai Bachchan (born Aishwarya Rai, and sometimes known simply as Ash or Gullu- November 1, 1973) is an Indian actress and former Miss World. Before starting her acting career, she worked as a model and gained fame after winning the Miss World title in 1994.

Aishwarya was born in Mangalore to Krishnaraj Rai and Vrinda Rai. She has one elder brother, Aditya Rai, who is an engineer in the merchant navy and has also co-produced one of Rai's movies, Dil Ka Rishta (2003). At an early age her parents moved to Mumbai where she attended the Arya Vidya Mandir high school in Santa Cruz. Aishwarya then entered Jai Hind College at Churchgate for one year, and then moved to Ruparel College in Matunga to finish her HSC studies. She did well in school and planned to become an architect and went on pursuing studies in architecture.

She can communicate in several languages, including her mother tongue Tulu, as well as Hindi, English, Marathi and Tamil. She started studying architecture but gave up her education to pursue a career in modelling. While pursuing

Fig- 6: A paragraph taken from internet

Count	Percentage	Found keyword
9	1.80 %	aishwarya
7	1.40 %	miss
7	1.40 %	bachchan
5	1.00 %	world
4	0.80 %	mumbai
4	0.80 %	actress
3	0.60 %	year
3	0.60 %	moved
3	0.60 %	abhishek
3	0.60 %	modelling
3	0.60 %	married
2	0.40 %	architecture
2	0.40 %	winning
2	0.40 %	studies

Fig- 7: Number of keywords, their relative frequencies and the keyword itself

b) Keywords Ranking: Important keywords are assumed to appear more frequently in a blog. Firstly, these keywords are ranked according to the number of appearances in a source. The more the number of appearances, the higher the rank. Once this has been done, the keywords having same frequencies are then re-ranked in accordance to their lengths. A longer keyword needs a bigger container, hence should be ranked accordingly.

3.1.3. Picture and Keywords Module:

A flowchart showing the basic steps in picture and keywords module is shown in Fig- 5(b).

(a) Patch vs. Keywords Correspondence: The aim of this step is that each and every keywords gets assigned to be filled in a container/patch which is capable of containing the keyword. Longer keywords should not be mapped to a smaller patch. Similarly, smaller keywords should not be put into larger patches. The keyword to patch correspondence should be performed in such a way that the keywords are put in appropriate containers, failing to do which may produce either overly sparse or overly

dense results which might be difficult to read and comprehend. Hence, sequential mapping of previously ranked keywords and patches should be done such that:

1) the important keywords have a more salient location, and 2) the longer keyword should be matched with a larger patch for better packing.

b) Warping Text to Image Patches: The most challenging part of this module is warping keywords into its corresponding patches. It is difficult because the patches are not all of the same size and shape, but vary according to the contour of the image boundary. Based on the correspondences between keywords and patches, the word is warped into the patches either manually using some image processing tools like Adobe Illustrator [37], or by using the algorithms proposed by various researchers. Hormann and Floater [33] presented the mean value coordinates for warping between two arbitrary planar polygons, based on a generalized notion of barycentric coordinates. In geometry, the barycentric coordinate system is a coordinate system in which the location of a point of a polygon (a triangle, tetrahedron, etc.) is specified as the center of mass, or barycenter, of masses placed at its vertices. In this work, warping is done manually [37].



Fig- 8: Image generated after warping text into relevant patches.

The output of the warping process is shown in Fig- 8. From the output of warping, one can easily make out various keywords that are used to recreate the original image. These keywords are easily comprehensible which adds to the information provided by the image.

3.1.4. Post-Processing Module: After looking at Fig- 8, one can easily make out the keywords related to the object of interest. But even after warping, there are lots of underutilized spaces in the image. This space is called blank space. For better visual effect, these blank spaces need to be filled. This can be done in several ways. The blank spaces can be flooded with special ASCII symbols (such as asterisk) for better visual effect. Regular text like

characters or numerals should not be used. This is done so that the viewers do not confuse these blank space fillers with the characters in the keywords. In this work, the output of the warping phase is merged with the original silhouette image of the object of interest. This fills up all the blank spaces within the patches in a better way than inserting special symbols, and enables better visibility of the PicWords. The final PicWords is shown in Fig- 9.



Fig- 9: Final output (PicWords) of the original image after post processing

In this current work, the final output can only be a binary image. Two other variations of PicWords, i.e., the Gray PicWords and the Colored PicWords have been implemented in [1]. It uses the binary PicWords as a mask to pick the gray/color value of the original image, respectively. Since image segmentation techniques are used, all the patches are spatially tightly connected and the fitted keywords are also tightly connected. Since all the patches are tightly connected, the keywords inserted into those patches are also close in proximity, hence it may be difficult for a human to differentiate between two adjoining keywords. Colored and grayscale PicWords can differentiate between different keywords more effectively than binary version. Some more results of the current work are shown in Fig- 10.



(a)



(b)



Fig- 10: Final outputs of PicWords system (a) a picture of Tom Cruise and its PicWords (b) a picture of Katrina Kaif and its PicWords

4. RESULTS AND DISCUSSION

The original PicWords uses Mean Shift segmentation technique for the initial segmentation of the image. Mean Shift is a powerful, yet versatile algorithm for performing image segmentation, node finding clustering etc. It is a non parametric iterative algorithm. The input to mean shift segmentation technique is a set of points or pixels from an image. If dense regions (or clusters) are present in the feature space, then they correspond to the mode (or local maxima) of the probability density function. For each pixel, Mean shift defines a window around it. The radius of the window is predefined and can vary according to the need of the user. For all the pixels lying in this window, mean shift computes the mean of the data points. Once the mean of these points (local maxima) is computed, the centre of the window is moved to this maxima. This process continues until the algorithm converges, and there is no more scope of moving the centre of the window to a new data point. In other words, after each iteration, the window moves to the data point corresponding to the densest region. Basically, mean shift works as follows:

1. A window is fixed around each data point. The window has a predefined radius, and all the data points lying within this radius are included in the window.
2. Calculate the mean of all the data points inside the radius of the window.
3. For each iteration, shift the window to the region of highest density.
4. Repeat steps 2 and 3 until the algorithm converges.

Mean shift algorithm takes quite some time. The time complexity of mean shift algorithm is given by ' $O(Tn^2)$ ' where 'T' is the number of iterations and 'n' is the number of data points (pixels in case of images) in the data set.

Fuzzy C Means (FCM), an unsupervised clustering and segmentation algorithm, is applicable to a variety of problems. In this approach, analysis is performed based on the distance between different input data points or pixels. Once the analysis begins, clusters are formed in

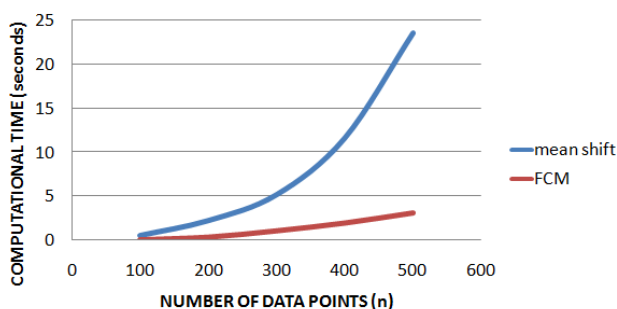
accordance to the distances between all the data points and each cluster has a central data point. In fact, in FCM, a set of data points or pixels are grouped together, such that each data point is related to every cluster. If it has a high degree of belonging to a certain cluster, then it will be placed in that cluster. However, if another data point is lying somewhere far away from the center of a cluster, then degree of belonging of that data point to that cluster will be relatively low. FCM iteratively moves the centers of the clusters to their correct location within a set of data points. To be very specific, once fuzzy logic in K-Means clustering algorithm is applied, and then it is called the Fuzzy C-Means algorithm. FCM clustering techniques are based on fuzzy nature of data points and they produce a clustering based on membership weights rather than some probabilistic functions. The behavior of this algorithm is similar to its supervised counterpart, i.e. the K-means clustering approach. The asymptotic efficiency of FCM can be expressed as $O(ndc^2i)$, where:

- i - Number of iterations of FCM over entire dataset.
- n - Number of available data points.
- c - Number of clusters required.
- d - Number of dimensions of the input.

Comparisons between the two methods was made on three sample pictures shown in Fig- 3(a) taken as sample 1, Fig- 10(a) taken as sample 2 and Fig- 10(b) taken as sample 3. The results are shown in Fig- 11.

Value of 'n'	Mean Shift (Time taken in seconds)	Fuzzy c means (Time taken in seconds)
100	0.527405	0.091743
200	2.234417	0.396979
300	5.181736	1.107128
400	11.64873	1.982176
500	23.6154	3.108042

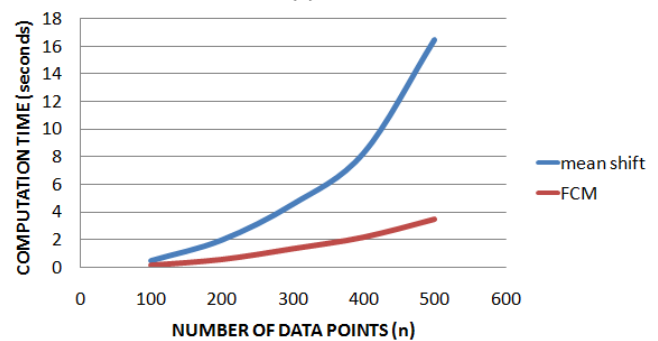
(a)



(b)

Value of 'n'	Mean Shift (Time taken in seconds)	Fuzzy c means (Time taken in seconds)
100	0.462759	0.116819
200	1.964181	0.538349
300	4.550385	1.317512
400	8.252429	2.155122
500	16.45374	3.460526

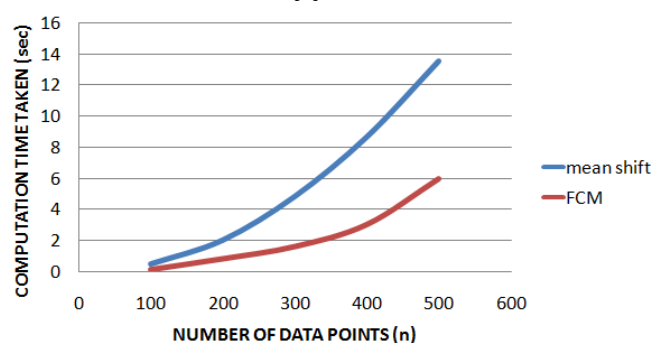
(c)



(d)

Value of 'n'	Mean Shift (Time taken in seconds)	Fuzzy c means (Time taken in seconds)
100	0.492806	0.149335
200	2.028282	0.849397
300	4.838397	1.635874
400	8.668452	3.053378
500	13.55449	6.007321

(e)



(f)

Fig- 11: Comparisons between mean shift method and fuzzy c-means method (a) and (b) for sample 1, (c) and (d) for sample 2, (e) and (f) for sample 3.

5. CONCLUSION AND FUTURE WORK

In previous work, the PicWords suffer from a few drawbacks. First of all, the original PicWords is deemed computationally expensive. This is because the segmentation of images is taking noticeable amount of time. The original PicWords uses Mean Shift segmentation technique for the initial segmentation of the image. Mean Shift is a powerful, yet versatile algorithm for performing image segmentation. However, the time complexity of mean shift algorithm is given by ' $O(Tn^2)$ ' where 'T' is the number of iterations and 'n' is the number of data points (pixels in case of images) in the data set. Also, the original PicWords adheres to the contour of the object of interest, along which keywords are placed. It is due to this property that images generated can resemble the original image while the regions filled by keywords contain informative text about the object of interest. But in case that the source image is textureless, the current system will probably fail.

To counter these issues, the use of a relatively cheaper segmentation technique, the Fuzzy c-means segmentation technique has been suggested. FCM clustering techniques are based on fuzzy nature of data points and they produce a clustering based on membership weights rather than some probabilistic functions. The asymptotic efficiency of FCM can be expressed as $O(ndc^2i)$, where:

i - number of iterations of FCM over entire dataset.

n - number of available data points.

c - number of clusters required.

d - number of dimensions of the input.

Furthermore, superpixel segmentation has been directly applied on silhouette images instead of the original image itself. A silhouette image is the dark shape and outline of someone or something visible in restricted light against a brighter background. Since silhouette images are textureless binary images, hence once a silhouette image of a textureless object is generated, it can be directly decomposed into patches which will give satisfactory results.

Of course, this current work is not perfect and has certain limitations. First of all, the various variants of PicWords, i.e. grayscale and colored PicWords are yet to be worked upon. Also, for generation of automated PicWords, warping of keywords into patches has to be performed by using the techniques suggested by other researchers based on barycentric coordinates. These limitations are scope of future work on the current system.

PicWords has great market potentials. It can be developed into an application for social networking websites to generate more informative profile pictures. It can be used in the field of advertisement, for example, it can be used in posters for promotion of various events, or advertisement of various products. The picture of the product can be used as source image, and it's features can be used as keywords to generate fancy PicWords.

REFERENCES

- [1] Z.Hu, S.Liu, J.Jiang, R.Hong, M.Wang and S.Yan, "PicWords: Render a picture by packing keywords", IEEE transactions on Multimedia, vol. 16, no. 4, June 2014, pp.1156 – 1164.
- [2] J. Xu and C. S. Kaplan, "Calligraphic packing," in Proc. Graphics Interface, 2007, pp. 43–50.
- [3] J. Kyprianidis, J. Collomosse, T. Wang, and T. Isenberg, "State of the „art“: A taxonomy of artistic stylization techniques for images and video," IEEE Trans. Visual. Comput. Graph., vol. 19, no. 5, pp. 866–885, 2013.
- [4] R. Carroll, A. Agarwala, and M. Agrawala, "Image warps for artistic perspective manipulation," ACM Trans. Graph., vol. 29, no. 4, pp. 127–135, 2010.
- [5] Haeberli, P., "Paint by Numbers". Proceeding SIGGRAPH '90, Proceedings of the 17th annual conference on Computer graphics and interactive techniques, pp. 207–214.
- [6] Dobashi J., Haga T., Johan H., Nishita T., "A method for creating Mosaic Images using Voronoi diagrams", in Proceedings of Eurographics 2002, September 2002.
- [7] Faustino G.M., De Figueiredo L.H., "Simple Adaptive Mosaic Effects", Proceedings of SIBGRAPI2005, August 2005.
- [8] Di Blasi G., Gallo G., "Artificial Mosaics", The Visual Computer, vol.21, issue 6, pp.373-383, 2005.
- [9] Battiato S., Di Blasi G., Farinella G.M., Gallo G., "A novel technique for mosaic rendering", WSCG 2006, January 2006.
- [10] A. Hausner, "Simulating decorative mosaics," in Proc. 28th Annu. Conf. Computer Graphics and Interactive Techniques, 2001, pp. 573–580.
- [11] Elber E., Wolberg G., "Rendering Traditional Mosaics", The Visual Computer, vol.19, issue 1, pp. 67–78, 2003.
- [12] Schlechtweg S., Germer T., Strothotte T., "Render Bots – Multi Agent systems for direct Image generation", Computer Graphics forum 24(2), pp.137-148, 2005.
- [13] Harmon L.D., "The recognition of faces", Scientific American vol.229 No.5, 1973.
- [14] Neret G., Descharnes R., Dali, "The Paintings", Taschen, Koln, 2001.
- [15] Close C., <http://www.chuckclose.coe.uh.edu/>, 1970
- [16] Silvers R., Hawley M., "Photomosaics", Henry Holt, New York, 1997.
- [17] Runaway Technology. <http://www.photomosaic.com/>, 2006.
- [18] hunt W.L., <http://home.earthlink.net/~wlhunt/>, 1998.
- [19] ArcSoft, PhotoMontage, <http://www.arcsoft.com/>, 2006.
- [20] Blake S., <http://www.barcodeart.com/>, 1998.
- [21] Di Blasi G., Petralia M., "Fast Photo Mosaic", Poster Proceedings of ACM/WSCG2005, January 2005.
- [22] Cantone C., Farro A., Pulvirenti A., Reforgiato Recupero D., Shasha D., "Antipole Tree indexing to support

range search and K-nearest neighbor search in metric spaces". IEEE/TKDE, vol.17, issue 4, pp.535-550, 2005.

[23] Strand C., "Hello, Fruit Face!: The paintings of Giuseppe Arcimboldo". Prestel, 1999.

[24] Di Blasi G., Gallo G., Petralia M., "Puzzle Image Mosaic" In proceedings of IASTED/VIIP2005, September 2005.

[25] J. Kim and F. Pellacini, "Jigsaw image mosaics," ACM Trans. Graph., vol. 21, no. 3, pp. 657-664, 2002.

[26] J. Orchard and C. S. Kaplan, "Cut-out image mosaics," in Proc. 6th Int. Symp. Non-Photorealistic Animation and rendering, 2008, pp. 79-87.

[27] X. Xu, L. Zhang, and T. T. Wong, "Structure-based ascii art," ACM Trans. Graph., vol. 29, no. 4, pp. 52:1-52:9, 2010.

[28] M. Nacenta, U. Hinrichs, and S. Carpendale, "Fatfonts: Combining the symbolic, and visual aspects of numbers," in Proc. Int. Working Conf. Advanced Visual Interfaces, 2012, pp. 407-414.

[29] R. Maharik, M. Bessmeltsev, A. Sheffer, A. Shamir, and N. Carr, "Digital micrography," in ACM Trans. Graph. (Proc. SIGGRAPH 2011), 2011, pp. 100:1-100:12.

[30] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 5, pp. 898-916, 2011.

[31] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 5, pp. 603-619, 2002.

[32] C. M. Christoudias, B. Georgescu, and P. Meer, "Synergism in low level vision," in Proc. 16th Int. Conf. Pattern Recognition, 2002, vol. 4, pp. 150-155.

[33] K. Hormann and M. S. Floater, "Mean value coordinates for arbitrary planar polygons," ACM Trans. Graph., vol. 25, no. 4, pp. 1424-1441, 2006.

[34] R. Achanta, A. Shaji, K. Smith, A. Lucchi and P. Fua, "SLIC Superpixels compared to State-of-the-art Superpixel Methods", IEEE Transactions on Pattern Analysis and Machine intelligence, col. 34, no. 11, November 2012, pp. 2274-2281.

[35] Xin-bo Zhang and Li Jiang, "An Image Segmentation Algorithm Based on Fuzzy C-Means Clustering", International Conference on Digital Image Processing ,Bangkok, 7-9 March 2009, pp. 22 - 26.

[36] Yin Li, Jian sun, Chi-keung Tang and Heung-Yeung Shum, "Lazy snapping", Proceeding SIGGRAPH'04, ACM SIGGRAPH 2004 papers, pp. 303-308.

[37] Adobe Illustrator, <https://www.adobe.com/in/products/illustrator.html>.

BIOGRAPHIES



Ankur Singh is currently a student of Birla Institute of Technology, Mesra, Ranchi. His area of interest is Image Processing.



Shruti Garg is currently an Associate professor at Birla Institute of Technology, Mesra, Ranchi. Her research interests are Image Processing and Soft Computing.