Accurate and Discernible Photocollages

Jordan Miller[1] and David Mould[1,2]

[1]University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[2]Carleton University, Ottawa, Ontario, Canada

Abstract
We propose a system for arranging images from a database into a collage that resembles some target image. These collages exploit large scale visual correspondences between the target image and the images in the database. We ensure that images of multiple sizes are used and are combined so that boundaries between images are not immediately apparently; as a result, the final collage consists of irregularly shaped image sections. The final collages contain a dynamic mixture of textures, images, and shapes that is in contrast to the geometric and regular character of many photomosaic techniques. In service of these tasks, we propose a fast scale-based method for querying an image library, a novel method for composing multiple images using geodesic distance Voronoi tesselations, and a novel base/detail method for shifting the colors of the final collage so that the target image is more accurately represented.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture and Image Generation—

1. Introduction

Photomosaics have been a powerful and fascinating tool for artists, combining multiple images into a single image that itself has a unique and coherent interpretation. Photomosaic creation involves arranging a set of images, usually on a regular grid or some other tessellation, so that the entire arrangement of images resembles some large target image. Correspondences between images may be used to add additional context and additional layers of meaning: it is possible to induce narrative effects such as irony, and to evoke viewer reactions such as surprise, confusion, or humor.

We propose an alternative form of assemblage which we term "photocollage", in which images are arranged irregularly, with content-sensitive boundaries, smooth blending between smooth image regions, and a high degree of both accuracy to the target image and discernibility of individual elemental images. As in photomosaics, individual images are combined in such a way as to suggest the features of the target image. However, we use larger images and recalculate regions of the collage where large images are insufficient to accurately represent the target image. Also, we seek large-scale structural matches between collage elements and the target image: in particular, strong edges in the target image should correspond with edges within collage elements, rather than boundaries between them.

The physical medium of collage in art uses an impressively diverse array of elements and composition techniques; element images vary widely in size, shape, and content [Hut68]. In particular, artists such as Dave McKean and Nick Bantock have explored a wide space of possibilities for expression and representation within the collage art form. Distinct from photomosaics, our system allows non-expert artists to compose textures and gradients that interact in visually complex ways, through smooth blending and content-sensitive seam finding. These photocollages consist of dynamic compositions of textures, colors and tones; the results are not the clean-cut, mechanical grids that we expect to see in photomosaics, but have a roughness and chaotic quality that brings them to life.

A photomosaic is an image with some large scale global interpretation composed of smaller images, each with their own local interpretation. Oliva and Torralba [OTS06] note that the property of an image having simultaneous global and local interpretations is distinct from the property of multiple global interpretations. Seckel [Sec04] uses the term *double image* to refer to any image that has two distinct global interpretations; that is, when viewed as a whole, these images

may be assigned two distinct interpretations by the viewer. For instance, some sixteenth century artists such as Matthaus Merian created *anthropomorphic landscapes* that at first appear to be simple landscape paintings but contain the hidden shape of a human face [Sec04] . However, since our photo-collages often use large images to represent large portions of the target image, it is useful to merge the ideas; that is, if we have a large portion of the target image represented by a single image in the collage, this collage exhibits weaker localization of images than the typical photomosaic.

After Silvers and Hawley published their picture book of photomosaics [SH97], Finkelstein and Range [FR98] published a formal treatment of the subject. Jacobs et al. [JFS95] and Zhang [Zha02] improved the efficiency of library queries by using content based image retrieval and low dimensionality signatures. Di Blasi et al. [DGP06] additionally extended the technique to create photomosaics with variable sized tiles according to a quad-tree partition of the target image. Orchard and Kaplan [OK08] extended the technique to handle arbitrary partitions and non-binary tile regions, and sped up the process of the exhaustive library search by using spectral techniques inspired by image registration methods. These techniques made feasible the use of optimal subimages from the image library or optimal rotations. The collages of Rother et al. [RBHB06] do not attempt to represent any secondary image, but images within each collage blend from one to the other without disturbing important image features. Similarly, the puzzle-like collages of Goferman et al. [GTZM10] pack irregularly shaped image regions into a collage so that they are arranged efficiently around each other. Kim et al. [KP02], Di Blasi et al. [DGM05], Gal et al. [GSP*07], and Huang et al. [HZZ11] all use individual object shapes rather than entire images as the matching primitives. Tong et al. [TZHM11] use edge-map matching and Poisson editing to hide secondary images in a primary image, similar to our method for post-processing. Correa and Ma [CM10] compose multiple scenes from a video sequence into an image which summarizes the narrative. Scenes are arranged linearly by time with continuous transitions, using graph cuts, from one scene to the next. Barnes et al. [BGSF10] accomplish a similar task, using an image retargeting method, but focus more on visual navigation through a video sequence.

Orchard and Kaplan identify the two primary, and sometimes conflicting, goals of the image mosaic process as *accuracy* and *discernibility*. *Accuracy* characterizes the efficacy of the mosaic in representing the target image; *discernibility* refers to the clarity and perceptual fidelity of the individual images within the collage. Orchard and Kaplan emphasize accuracy by using fragments of each image in the mosaic, rather than the complete images, thereby increasing accuracy. Orchard and Kaplan argue that more accurate matches allow them to use larger tiles, thereby increasing discernibility. However, their use of incomplete and fragmentary subimages simultaneously reduces discernibility.

We seek to achieve simultaneously high accuracy and high discernibility. Like Orchard and Kaplan, we also use a subregion of each image rather than the entire image. We adjust the balance between discernibility and accuracy as a function of position within the image. Some regions of the image do not require as much accuracy as others, and thus we are free to focus on discernibility in these regions. For example, an image region containing something like an empty blue sky, devoid of detail, does not demand a close match; rather, we may quickly find an approximate image match in the library and represent it as discernibly as possible.

To find the best region of an image to use for a given section of the collage, we propose a fast hierarchical matching process in which several matching passes are performed, each with a more detailed set of features and each within a reduced image library. Methods of signature matching, as demonstrated by used by Di Blasi et al. [DGP06], while extremely fast, have not been extended to handle subregions of images, and can be inaccurate in some cases. Orchard and Kaplan recognize these shortcomings and provide a solution; however, their method is still time-consuming. We have achieved competitive speed by exploiting the multi-scale nature of images to find high quality matches between subimage regions.

Finally, we propose a novel color correction scheme. Orchard and Kaplan proposed a tile-wise color correction scheme, shifting luminance and chrominance uniformly across each tile to better match the target image. This method sometimes shifts the colors in unconvincing ways, producing improbable colorings such as the orange skies seen in figure 5 (or blue cows, visible here if one looks closely). The Jigsaw Image Mosaic and Puzzle Image Mosaic systems use color correction on the object level, shifting the colors of each object towards the colors of the target image. This approach is appealing for our application, but since we perform matching on the image level, we find a way to approximately isolate image objects, by using a base and detail layer decomposition [BPD06, FFLS08]. Even when applied to arbitrary textures, not necessarily photocollages, our color correction can create fascinating double images.

2. Algorithm

The inputs for the algorithm are: a *target image*, to be recreated as a collage; an *image library* consisting of many images to be used as collage elements; and values for various parameters. For the target image, any image may be used but we have found that results are more attractive for high contrast images with a few clearly discernible objects. Likewise, most images are suitable to be included in the library, although we have found that images with more interesting, dynamic content provide more visually appealing results.

Our algorithm for synthesizing image collages uses six main steps:

1. **Preprocessing** We preprocess the library images to obtain data used to accelerate and inform match calculations. The contrast of the target image may be increased to increase the likelihood of interesting matches.

2. **Partitioning** The target image is partitioned into tiles via a centroidal Voronoi diagram process. Several partitions of varying coarseness are computed.

3. **Matching** Each tile is matched with a fragment of some image from the library. Matching is performed in stages, from coarse to fine, paring down the size of the library with each pass until a final match is selected. For each image, multiple comparisons are made between the image and the tile, one for each valid translation of the image against the tile.

4. **Compositing** Once an image fragment has been chosen for each tile, the Voronoi partition is recomputed using the same centroids, but with geodesic distance rather than Euclidean distance, so that the content of the images can adjust the shape of the seams between images.

5. **Error Accounting** The error density within each tile in a finer partition is calculated, and the tiles with the highest error density are recomputed.

6. **Color Correction** Using a base/detail approach, the target image and initial collage are combined into a final collage which is even more accurate but still contains clearly discernible collage images without the appearance of any imposition of the raw target image on the collage.

We next discuss each step at greater length.

2.1. Preprocessing

We preprocess the library images to extract histograms and a pyramid of weight maps. The histograms are used as signatures for an initial image rejection step and the weight maps are used to refine later match calculations. We also generate downsampled versions of each library image at several resolutions.

The purpose of the weight maps is to inform match calculations with local features. The critical features are the intensity edges. Counterintuitively, we recommend assigning near-zero weight very close to an edge, then a high weight at a slightly greater distance, and low but nonzero weights far from the edges. Our suggested shape for the weight function F is illustrated in Figure 1.

A weight map with this shape allows us to balance between accuracy and discernibility. Given a finite library, we are unlikely to be able to match exactly all edges: a mismatch near an intensity edge is likely to have quite high error. We plan to subdivide tiles that contain excessive error, but we prefer to use large tiles rather than always split tiles near edges. Relaxing our error tolerance on edges will allow us to use larger tiles, hence promoting discernibility.

The weight map F is based primarily upon a large-scale gradient magnitude edge map $F_1 = \nabla G_r(T)$ (where $G_R(.)$ indicates a Gaussian convolution of radius R), with some modifications. Particularly, we invert the gradient magnitude edge map $(1 - F_1)$ so that non-edge regions have value 1 and edge regions fall toward 0. In order to enforce the falloff of the map toward some small value c in regions far away from edges, we blur the gradient magnitude edge map F_1 by some large radius $R > r$ to obtain $F_2 = G_R(|\nabla G_r(T)|)$. By multiplying F_1 and F_2 and ensuring a minimum value c in empty regions, we obtain a map F with the desired properties:

$$F \;=\; (1 - F_1)\Big[F_2(1 - c) + c\Big], \text{ or} \qquad (1)$$

$$F \;=\; |1 - \nabla G_r(T)|\Big[G_R(|\nabla G_r(T)|)(1 - c) + c\Big], \quad (2)$$

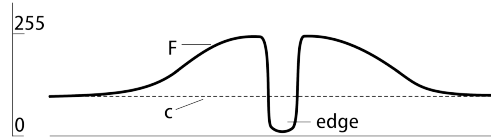where T is the target image and $G_R(.)$ indicates a Gaussian convolution of radius $R > r$.

Figure 1: A one-dimensional cross-section of F. The map falls to a neutral value c in empty regions of the image.

2.2. Partitioning

We next partition the target image into a set of tiles; each tile will eventually contain some image from the library. We intend to place tiles so that important image edges are contained therein, attempting to promote interesting correspondences between the image objects and the portions of the target image they represent.

We use a centroidal Voronoi tesselation with centroids generated randomly. In practice, we calculate a sequence of tesselations, starting with a coarse tesselation with few tiles; each subsequent tesselation in the sequence has a greater number of smaller tiles. With a typical image library, accurate collages generally cannot be made using coarse tilings. However, the inaccuracies in such a collage are often very localized, in which case the inaccurate regions can be recalculated using a finer partition. For the density function, we use a coarse gradient magnitude map from the target image. The overall process produces tiles that are centered on strong edges in the target image, convex in shape, and roughly uniform in size.

2.3. Matching

Having obtained a partition of the target image, we match each tile in the partition with an image from the library. For

each tile, we perform a sequence of evaluation passes to determine match fidelity between the tile (or the target image therein) and some set of images in the library. After each evaluation pass, we reduce the library to some subset of images likely to be good matches, until, after the final step in the sequence, the library contains just one matching image. Note that this process begins with a large number of low-resolution, fast comparisons but progresses in later iterations to a small number of detailed comparisons.

In the first pass, we compute a coarse histogram for each image in the library and discard the images that have an insufficient number of pixels in the bin corresponding to the mean luminosity of the current tile. We then compare every translation of every image in the reduced library with each tile. This comparison pass is repeated several times at different scales, increasing the resolution with each pass. Typically, two scales are sufficient; in this case, we perform a pixel-wise comparison of each image at some low resolution, reduce the library to the best matches from that pass, and then repeat the process at a higher resolution to choose the final match.

For each translation of each image in the library we evaluate the quality of the match with the tile by summing the match score of each pair of pixels between the image and tile. We evaluate the absolute difference between the luminosity and chrominance values of the pixels, and multiply this value by a feature map. For each library image I, for each translation, and for a given channel $n = \{l, c\}$, representing luminance and chrominance, a match score is computed in the following way:

$$E_n = \sum_{(x,y) \in A} F(x,y)M(x,y)|I_n(x,y) - T_n(x,y)|, \quad (3)$$

where $A = I \cap T$ is the rectangular region of intersection between image I and target image T. F is the weight map from section 2.1. The function $M \in [0,1]$ represents the tile mask, where $M(x,y) = 1$ indicates that the pixel (x,y) is inside the tile, $M(x,y) = 0$ indicates that the pixel (x,y) is outside the tile.

For an image I and a given translation, the final ranking is then determined from the following quantity:

$$E = w_l E_l + w_c E_c, \quad (4)$$

where $\vec{w} = (w_l, w_c)$ are parameters chosen by the user to influence the fidelity of the collage in matching the colors of the target image. We weight the luminance term w_l higher than the chrominance term w_c to improve perceptual accuracy.

Before comparing each image with the current tile, we first rescale it to some proportion of the largest dimension of the tile. A scaling factor of 1 indicates that the tile just fits inside the image without room for translation. A larger scaling factor indicates that the image is larger than the tile and thus several options exist for matching a subimage of the current

image with the tile. Larger scaling factors may result in more accurate matches at the expense of longer processing times and less discernible collage images.

2.4. Compositing

We wish to fit the images in the collage together so that important image content is preserved and boundaries between images do not cut through important edges. To this end we compute another Voronoi partition with the original centroids, but using a geodesic distance rather than a Euclidean metric [CSRP10]. The geodesic weight of each pixel is determined by taking the maximum of all gradient magnitude edges of all collage images overlapping at that pixel. (note that gradient magnitude edges are weighted to fall off as a function of distance from the boundaries of the *original* Voronoi tiles–this is to ensure that the new tile boundaries do not deviate too far from the old ones.) More formally:

$$e_{\text{cost}}(T,x,y) = \text{Max}(\{|\nabla(I_i(x,y))| \cdot v_i(x,y)\}), \quad (5)$$

for all library images I_i and where $v_i(x,y) = G_R(\text{TileBoundary}_i(x,y)) \in [0,1]$ is proportional to the distance from a boundary in the original Voronoi tessellation, with greater weight closer to the boundary. Note that $\text{TileBoundary}(x,y) \in [0,1]$ is a binary map with tile boundaries $= 1$, and R is a size parameter.

When compositing the final collage, boundaries between images are feathered with width inversely proportional to the strength of the image edges at the current position. A narrow feathering preserves hard edges when needed, and a wide feathering smoothly transitions between smooth image regions.

2.5. Iteration

As mentioned in section 2.2, we iterate the photocollage process in order to increase accuracy where necessary. Wherever the accuracy of the collage is insufficient after iteration n, we recalculate certain tiles in iteration $n + 1$. The newly calculated tiles are smaller than the tiles in the previous iteration, to facilitate accurate matching. We recalculate only those areas of the image that have the highest error; we use the L1 norm for error, but L2 or other error metrics are possible.

Following collage composition, an error density is calculated for each tile in the tessellation for the next refinement pass. The tiles with the highest error density are activated for processing in the next iteration. For most of the results we report, we recalculate a fixed proportion of 33% and 25% of the total number of tiles for the second and third passes respectively, or approximately 50% when one refinement pass is used.

2.6. Color Correction

After all iterations have been completed, we seek to increase the accuracy of the collage by doing a final color correction pass. For the examples in this paper, we reduce processing time by processing each collage at half resolution and then reintroducing high frequency data from the original collage. We create a base layer P_B that contains large scale features from the target image and compose it with a higher-frequency detail layer P_D that contains details from the collage. Additionally, we compose another detail layer P_0 to represent the details from the target image that may have been poorly represented by the collage process.

We create the base layer P_B by performing a cross-bilateral filter on the target image T using the collage as the cross image. More specifically, we use a version of the collage image that has had textures removed through some edge-preserving abstraction operation, for we wish to create a base layer without excess detail.

$$P_B = B_{\text{cross}}(T, C_B, R), \qquad (6)$$

where $B_{\text{cross}}(.,.,R)$ represents the cross-bilateral filter of radius R and $C_B = \text{Abstract}(C, R')$ represents an abstracted version of the collage C computed with some scale R', possibly distinct from R. We choose R to be approximately the scale of the collage tiles or image objects.

For the function $\text{Abstract}(\cdot)$, any existing abstraction process could be used. We propose our own abstraction method in section 2.7.

The process of equation 6 effectively blurs the target image everywhere except across boundaries between objects and edges in the collage image, resulting in object-coherence within the base layer.

The base and detail layers are combined as follows:

$$\begin{aligned} P &= P_B + (C - \text{Abstract}(C, R')) & (7) \\ &= P_B + P_D. & (8) \end{aligned}$$

An overview of the color correction process so far is shown in Figure 2.

This process may still omit some salient details from the target image. To capture these, we create another detail layer P_0 by first repeating the process used to obtain P_B, but with a smaller filter radius r (proportional to the scale of the details we wish to represent). We then take a Gaussian highpass of the result to extract the details. We want to avoid adding extra detail to smooth regions of the collage, so we multiply the result by a map τ representing the contrast of textures in the collage image. In section 2.7 we describe the texture contrast map τ in more detail. Additionally, it is useful to multiply the result by a large scale gradient magnitude map D from the

target image to prevent the over-emphasis of edges:

$$P_0 = D \cdot \tau(C) \left[B_{\text{cross}}(T, C, r) - G_r(T) \right]. \qquad (9)$$

The final collage is then expressed as follows:
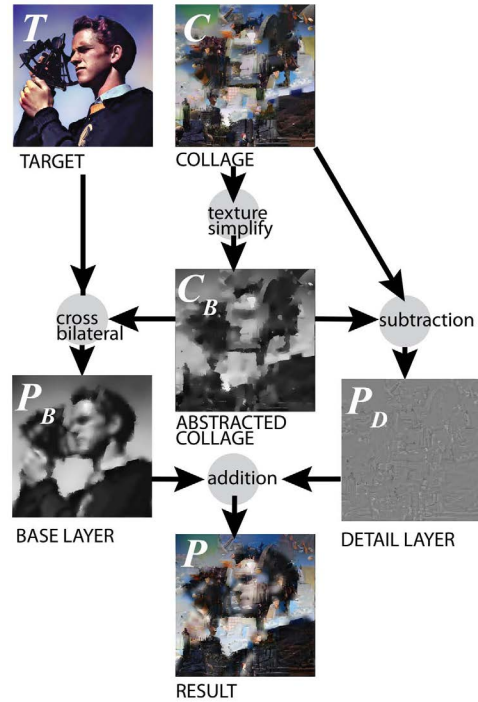
$$P_{\text{final}} = P_B + P_D + P_0. \qquad (10)$$

Figure 2: A schematic diagram of the color correction process, before the detail layer P_0 is applied. Notice that the base layer P_B is informed by both the target image and the collage image.

2.7. Image Abstraction

Many edge-preserving smoothing and image abstraction processes have appeared in the literature [TM98, FAR07, FFLS08, CM02, OBBT07]. For the function $\text{Abstract}(\cdot)$, we propose a bilateral filter with the intensity distance modulated by a texture contrast map $\tau(I)$; that is, for pixels with low texture contrast, weak filtering occurs, whereas for pixels with high texture contrast aggressive filtering occurs. This modified bilateral filter smooths even high contrast textures while preserving edges. Conceptually, this method is similar to the work of Su et al. [SDA05] in de-emphasizing regions of high texture activity.

For the texture contrast map $\tau(I)$ we use an operation inspired by the *textureness* map of Bae et al. [BPD06]. Bae
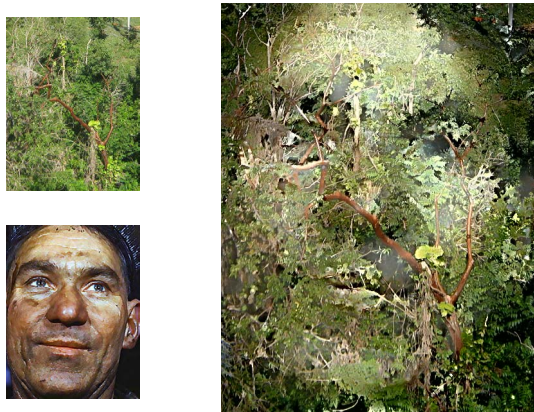
Figure 3: A face image is combined with a highly textured image.

et al. measure the textureness of an image I by taking a highpass filter $H(I)$ and then filtering its magnitude $|H|$ using a cross-bilateral filter against I. However, this textureness map spuriously highlights edges, which should be distinct from textures, and shows poor homogeneity in certain texture regions. As the contrast of the textures being measured increases, the homogeneity of the textureness map decreases, since the bilateral filter becomes less effective in smoothing. However, one can construct a similar filter by replacing the Gaussian highpass filter $|H| = |I - G(I)|$ with the filter $|H_m| = |I - M(I)|$ (where $M(I)$ is the median filter of I) and completing the operation with a median filter rather than a cross-bilateral filter. Thus Textureness $= \tau(I) = M(|I - M(I)|)$. The resulting texture contrast map is smoother and is not contaminated by edges; the median filter suppresses line edges and smooths out textures regardless of contrast. The disadvantage is that some spatial precision is lost, due to the propensity of the median filter to obliterate corners and other fine-scale features.

We have observed that with repeated iteration of this textureness process, the magnitude of the resulting map decreases. This property allows us to create a series of textureness maps, with decreasing magnitude but *increasing* precision (decreasing filter radius). Each term in this sum, being more precise than the previous, fills in the attenuated details from the previous term; in this way, we are able to create a textureness map $\tau(I)$ that is smooth, respects image edges and corners, and is spatially precise. Figure 4 shows two terms in this series and the final result, as well as a comparison of our textureness map with the methods of Bae et al. and Carson et al. [CBGM99].

We have also observed that performing N iterations of the function $I_n = |I_{n-1} - M(I_{n-1})|$ where $I_0 =$ the original image, the resulting map seems to highlight salient features

such as edges and corners while suppressing smooth or regularly textured features (we use N=7). We have used this iterated process to remove non-texture edges from our weight map F, so that textured regions are not falsely identified as belonging to object edges.

3. Evaluation and Conclusion

Accuracy. Figure 5 shows a comparison between a collage created through our method and a mosaic from Orchard and Kaplan [OK08]. The contrast of the target image was increased to promote interesting matches and the original contrast was reapplied in colour correction. We compared the accuracy of these collages in representing the original target image by computing the average per-pixel Structural Similarity value for the luminosity channel. Following Wang et al. [WSB03], we computed the SSIM measurement at several scales. While our results have a lower MSE value over all scales (lower MSE indicating a better match) and higher structural similarity at coarse scales (higher SSIM indicating a better match), at fine scales the results from Orchard and Kaplan have better SSIM scores.

Table 1: MSE and SSIM

MSE				
scale (finest to coarsest)	1	2	3	4
Our result	432	330	242	144
Orchard and Kaplan	540	499	435	336
SSIM				
scale (finest to coarsest)	1	2	3	4
Our result	0.41	0.45	0.57	0.71
Orchard and Kaplan	0.45	0.43	0.49	0.62

Overall, we can say that these two results have similar accuracy, especially when also considering visual inspection and visual comparison with the target image. It may even be argued that our result has higher accuracy when considering the fidelity of the reproduction of the salient features of the farmer's face.

Discernibility. Figure 5 shows a detailed comparison of several sections from the collage and mosaic, for the purpose of judging discernibility. In figure 5f we see that a house and a family are immediately discernible in our result and in figure 5e we immediately discern a map and train yard; in the previous results more prolonged inspection is required to discern comparable details. Also, rather than filling empty areas of the image with many empty photographs, our algorithm has chosen a few large representative images as we see in figures 5d and 5e. Figure 6 shows some especially discernible examples. In figure 6a we see clearly in the detailed section a sailor (top) and welder (bottom) and in figure 6b we see clearly a woman in a white shirt (top) and a man in a hat looking toward the camera (bottom). Also notice the edge
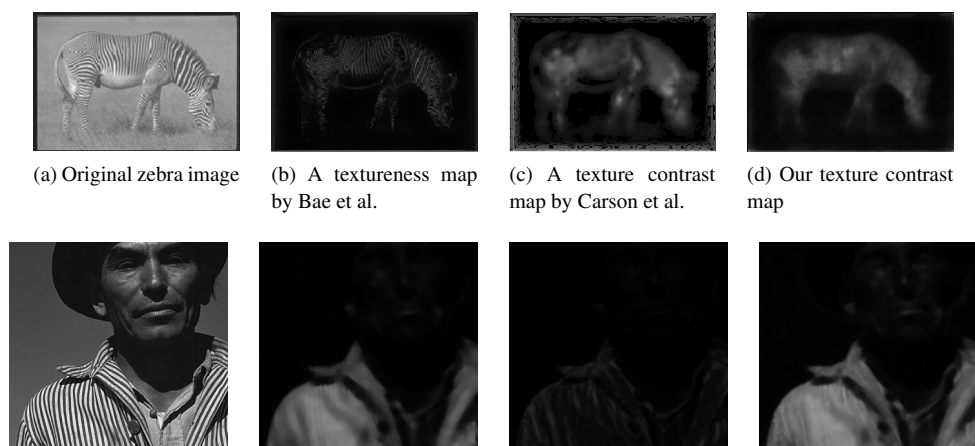
(a) Original zebra image   (b) A textureness map by Bae et al.   (c) A texture contrast map by Carson et al.   (d) Our texture contrast map

Figure 4: *Each successive term increases in detail but decreases in magnitude. The rightmost image shows the sum of the terms.*

matching: for example, the woman's white shirt approximately matches the edge of the pilot's headgear silhouetted against the sky.

Color Correction. As with the method of Orchard and Kaplan, our color correction algorithm dramatically increases the accuracy of a collage. Even a collage that initially exhibits very poor accuracy may be repaired in this way. However, we have been able to avoid the appearance of strangely colored objects and distracting false details from the target image. Figure 5 shows an overall comparison of our method with that of Orchard and Kaplan.

Parameters. The large parameter space of our system may lead to difficulties in finding a set of parameters that produce good results for a large set of inputs. We provide recommended parameter settings, empirically generated, suitable for an image of size approximately 1000×1000. We suggest two or three levels of refinement, with each level containing between 3 and 6 times more tiles than the previous level. The radius of the coarsest tiles is approximately 150 pixels. We weight the influence of luminance approximately 3 times greater than chrominance in matching: $(w_l, w_c) = (1, 0.3)$. For color correction, we use a filter radius of about the same radius as the smallest tiles, or approximately 25 pixels, potentially varied by up to a factor of 2. For the scale factor S, we recommend a value of 1.3 for higher discernibility or 2.0 for higher accuracy.

Timing and Efficiency. For a target image of size 1400×945 pixels, with a database of 79 images, each of size $655 \times 500 \pm 50$ pixels, we measured a time of 17.0 minutes for determining all matches not including the calculation of edge maps, compositing, and color correction. The preprocessing step took a total of 50 seconds for all images. Calculating the time for two more databases (of size 53 and 27 images) we found that the average rate of matching was 13 seconds per

file. The per-image rate of matching should be sub-linear for an optimized implementation. This test was conducted with two additional refinement passes after the first and a uniform scaling factor of $S = 2.0$. The proportion of tiles calculated was 33% and 25% for each of the two refinement passes respectively. The computer used to perform these tests was a Dell XPS 420, with the Intel Core 2 Duo 3.00Ghz processor and 3.00Gb of RAM. Our implementation has been programmed in Java 6.0.

Limitations. Our refinement scheme, which processes new tiles based on error density, often damages the discernibility of the collage image by partially covering previously discernible images. Figure 7 shows the success of a collage in which most regions of refinement were hand-picked by a user. While using our system in an artistic context we have noticed that it is often necessary to manually remove the effects of color correction from faces, indicating that face detection may be useful, as in Autocollage.

Sometimes the color correction will reverse dark/light relationships between image objects in order to better mimic the target image, damaging discernibility. Techniques to mitigate this effect have been proposed by others, particularly Bae et al. [BPD06].

We consider repetition of images to be aesthetically damaging. In our system tile size variability ameliorates this somewhat, but it still may pose a problem in some cases. Particularly, for a target image that is dominated by a narrow band of tones, there may not be a wide array of choices within the library to represent these tones. This could possibly be mitigated by some policy for histogram equalization and local contrast adjustment in the target image. Similarly, for image libraries that contain a large number of images with similar mean luminance, the algorithm may choose the same image repeatedly. Through experience we have found that libraries
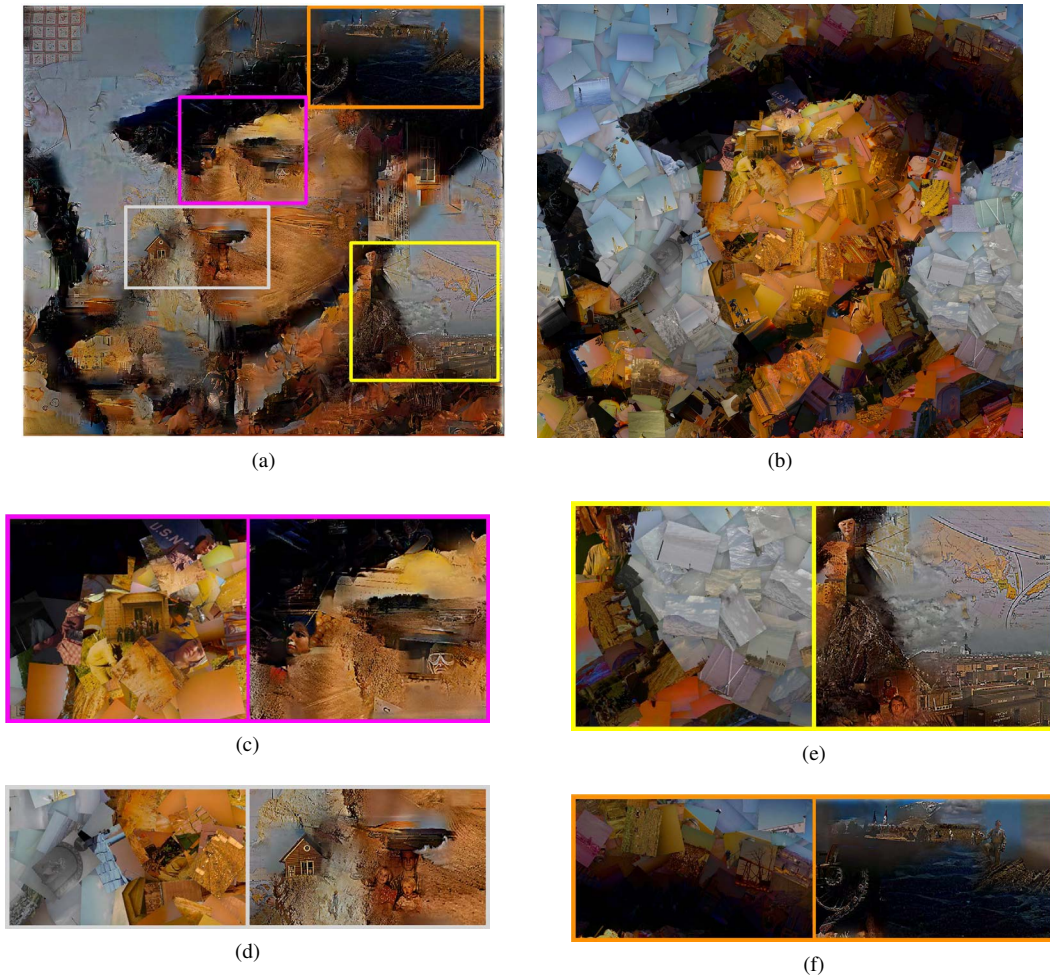
Figure 5: A detailed comparison of our collage with Orchard and Kaplan's photo mosaic. Each detailed section corresponds to the section of the target image with the identically colored border.

often lack sufficient numbers of very light and very dark images. Techniques to mitigate this include some randomization of choices or management of the mean luminance of the library images in a library-wise fashion.

4. Conclusion

We have described a system to produce a collage with an accurate global interpretation and discernible local interpretation using large images that smoothly flow into one another. Additionally, our system produces collages that are richly textured, consisting of images varying widely in shape and size. In figure 7 we show that a user may assist in the selection of collage elements in order to guide the process to a more attractive result. In figure 6 we show conversely that an automatic result may also be sufficient to represent the target image. Our system is able to produce a collage of accuracy approximately equal to that produced by previous systems in a short amount of time using fewer and larger component images with higher discernibility. Furthermore, the seamless composition of images and use of interesting large scale correspondences between images lends itself to an entirely different and broader style of artistic creation. We hope that artists will be able to use this system to increase their productivity and experiment with new ideas regarding double images.
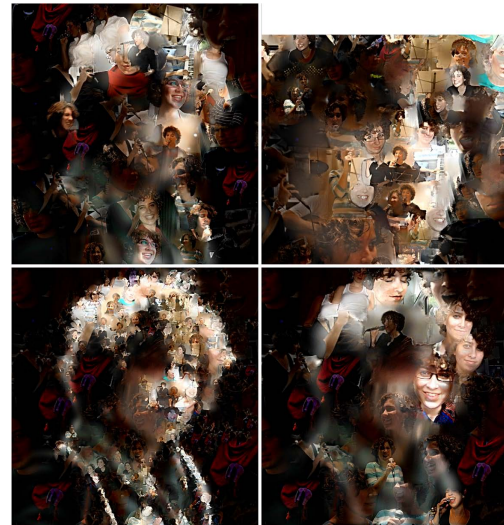
(a)

(b)

Figure 6: Two photocollage examples demonstrating high discernibility. Both collages have been tuned to favor the chrominance of the target image.

(a)

(b)

Figure 7: This collage (figure 7b) has been created by an artist from four image collages (figure 7a), each created with different parameters. The artist has erased the detail enhancement layer near faces, in order to improve discernibility, and applied some bilateral filter-based detail enhancement. Also, chrominance from the target image has been blended into the final collage to improve perceptual accuracy.

References

[BGSF10] BARNES C., GOLDMAN D. B., SHECHTMAN E., FINKELSTEIN A.: Video tapestries with continuous temporal zoom. In *ACM SIGGRAPH 2010 papers* (New York, NY, USA, 2010), SIGGRAPH '10, ACM, pp. 89:1–89:9. 2

[BPD06] BAE S., PARIS S., DURAND F.: Two-scale tone management for photographic look. *ACM Trans. Graph. 25*, 3 (2006), 637–645. 2, 5, 7

[CBGM99] CARSON C., BELONGIE S., GREENSPAN H., MALIK J.: Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24* (1999), 1026–1038. 6

[CM02] COMANICIU D., MEER P.: Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 5 (2002), 603–619. 5

[CM10] CORREA C. D., MA K.-L.: Dynamic video narratives. In *ACM SIGGRAPH 2010 papers* (New York, NY, USA, 2010), SIGGRAPH '10, ACM, pp. 88:1–88:9. 2

[CSRP10] CRIMINISI A., SHARP T., ROTHER C., PÉREZ P.: Geodesic image and video editing. *ACM Trans. Graph. 29*, 5 (Nov. 2010), 134:1–134:15. 4

[DGM05] DI BLASI G., GALLO G., MARIA P.: Puzzle image mosaic. In *IASTED/VIIP2005* (2005). 2

[DGP06] DI BLASI G., GALLO G., PETRALIA M. P.: Smart ideas for photomosaic rendering. In *Eurographics Italian Chapter Conference* (2006), Gallo G., Battiato S., Stanco F., (Eds.). 2

[FAR07] FATTAL R., AGRAWALA M., RUSINKIEWICZ S.: Multiscale shape and detail enhancement from multi-light image collections. In *SIGGRAPH '07: ACM ACM Transactions on Graphics (TOG) 2007 papers* (New York, NY, USA, 2007), ACM, p. 51. 5

[FFLS08] FARBMAN Z., FATTAL R., LISCHINSKI D., SZELISKI R.: Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph. 27*, 3 (2008), 1–10. 2, 5

[FR98] FINKELSTEIN A., RANGE M.: Image mosaics. In *Electronic Publishing, Artistic Imaging and Digital Typography, Lecture Notes in Computer Science Series, number 1375* (1998), Hersch R. D., André J., Brown H., (Eds.), Springer-Verlag, pp. 11–22. 2

[GSP*07] GAL R., SORKINE O., POPA T., SHEFFER A., COHEN-OR D.: 3D collage: expressive non-realistic modeling. In *NPAR '07* (New York, NY, USA, 2007), ACM, pp. 7–14. 2

[GTZM10] GOFERMAN S., TAL A., ZELNIK-MANOR L.: Puzzle-like collage. *Computer Graphics Forum 29*, 2 (2010), 459 – 468. 2

[Hut68] HUTTON H.: *The Technique of Collage*. B.T. Batsford Ltd, Watson-Guptill Publications, 1968, ch. Part one: Techniques, pp. 11–13. 1

[HZZ11] HUANG H., ZHANG L., ZHANG H.-C.: Arcimboldo-like collage using internet images. In *2011 SIGGRAPH Asia Conference* (New York, NY, USA, 2011), SA '11, ACM, pp. 155:1–155:8. 2

[JFS95] JACOBS C. E., FINKELSTEIN A., SALESIN D. H.: Fast multiresolution image querying. In *SIGGRAPH '95* (New York, NY, USA, 1995), ACM, pp. 277–286. 2

[KP02] KIM J., PELLACINI F.: Jigsaw image mosaics. *ACM Trans. Graph. 21*, 3 (2002), 657–664. 2

[OBBT07] ORZAN A., BOUSSEAU A., BARLA P., THOLLOT J.: Structure-preserving manipulation of photographs. In *NPAR '07* (New York, NY, USA, 2007), ACM, pp. 103–110. 5

[OK08] ORCHARD J., KAPLAN C. S.: Cut-out image mosaics. In *NPAR '08* (New York, NY, USA, 2008), ACM, pp. 79–87. 2, 6

[OTS06] OLIVA A., TORRALBA A., SCHYNS P. G.: Hybrid images. *ACM Trans. Graph. 25*, 3 (2006), 527–532. 1

[RBHB06] ROTHER C., BORDEAUX L., HAMADI Y., BLAKE A.: Autocollage. *ACM Trans. Graph. 25*, 3 (2006), 847–852. 2

[SDA05] SU S. L., DURAND F., AGRAWALA M.: De-emphasis of distracting image regions using texture power maps. In *APGV '05* (New York, NY, USA, 2005), ACM, pp. 164–164. 5

[Sec04] SECKEL A.: *Masters of Deception: Escher, Dali & Artists of Optical Illusion*. Sterling Publishing Co., Inc., 387 Park Avenue South, New York, NY 10016, 2004. 1, 2

[SH97] SILVERS R., HAWLEY M.: *Photomosaics*. Henry Holt and Co., 1997. 2

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on* (Jan 1998), pp. 839–846. 5

[TZHM11] TONG Q., ZHANG S.-H., HU S.-M., MARTIN R. R.: Hidden images. In *NPAR 2011* (New York, NY, USA, 2011), ACM, pp. 27–34. 2

[WSB03] WANG Z., SIMONCELLI E. P., BOVIK A. C.: Multiscale structural similarity for image quality assessment. In *in Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers, (Asilomar* (2003), pp. 1398–1402. 6

[Zha02] ZHANG Y.: *On the use of CBIR in image mosaic generation*. Tech. Rep. TR 02-17, Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, 2002. 2