Off you go!

Sai Chowdary G
CS 603
Computational Photography
Indian Institute of Technology Gandhinagar

***Abstract*—The popularity of Digital photography has been rising rapidly in the recent times. A large number of photographs being taken everyday, everything from special events to normal day to day activities are being captured. The proliferation of smart phones with quality inbuilt cameras has especially helped this observed trend. The most commonly captured scenes are those of monuments, natural scenes, parties or indoor activities etc by tourists, guests etc. But in most of these images due to the nature of the scene being captured the object of interest often gets occluded, for example the monument being captured might get occluded by other tourists who are moving about. In such cases one can capture multiple images of the scene such that the information lost in one image of the object of interest is available in the other image. In this project an attempt is made to remove such occluding objects and inpainting the image with the information of interest from the other image of the same scene using the principles of Homography, Laplacian Blending, Seam Carving etc.**

I. Introduction

Of late due to the technological advancements it has become possible to have good quality inbuilt digital cameras because of this there has been a large increase in the number of digital photographs being captured. One of the common problem that vexes people is that of occlusion i.e the object of interest in the scene is often occluded (covered) by other objects in the scene or other people etc. If it is some monument or a famous building, chances are that you will not be able to capture it completely as almost surely there will be other tourists around you who are trying to capture the building from different views and it would be a herculean task to get all of them out of your frame of view so how much ever we try these people will end up occluding some part of the building or the other. Or in some other cases we might want to be the only person in the photo like at a party scene etc but here too some other people might come invariably be captured into the image along with you (though this can not exactly be called as occlusion, we would still prefer to not have such objects in the image). In some other case, from an old photograph of some house etc we might want to remove some objects or people occluding it in the image so that we have the complete place in the captured. From the above scenarios it is clear that there are a large number of photographs from which we would like to remove the occluding objects and try to fill up that patch by what should actually have been there had the occluding object not been present while capturing the image. One simple solution would to crop the object from the scene if possible, or even better remove the object using seam carving[], while this solution does succeed in removing the unwanted objects it does not give the original information of the object of interest that we want, it just throws away the unwanted stuff, moreover in some cases this can distort the object of interest as well (this can happen when the amount of occlusion is high). To address this issue we need to some how get the actual information of the occluded part, this can indeed be obtained from another image of the same scene but taken from a different view(it is also possible to get the lost information from an image taken from the same view but a after some time as the object if it is a person might move to a different point and will not be occluding the previous patch now). Then we must replace the occluding object with this information by inpainting. We use Homography to geometrically align the images, then we find the patch to be replace the occluding object using the computed Homography and then combine this patch with the original image using laplacian blending. Also in case the images are capture by different cameras or lighting conditions this would make the final result less appealing so to account for this we come up with a multiplication factor to adjust the luminance values.

II. Related Works

This implementation follows the rough outline of the technique suggested by [[**?**]] though many changes have been made to it. Previous works of image Inpainting vary a lot in their efficiency depending on various factors like the surroundings of the region to be filled, and the size of the region to be filled. For a few algorithms[[1], [2]] if the size of the region is small enough then using the surrounding information we can fill the region but this will fail in the case of larger regions that need to be filled. To account for this new algorithms[[1], [3], [4]] have been proposed which take into consideration the salient structures and textures across the region. But still they do not perform to our satisfaction as in all these cases we only try to predict the region to be filled from the one occluded image that we possess. A few algorithms[[5], [6], [7]] use the images of the same scene but under the constraint that the images be captured by the same camera. Using this technique we can fill regions of almost any size and also the images need not be of the same camera as long the object of interest is almost planar(minor depth variations will not be an issue).

III. Proposed Approach

Two images are sufficient to implement the proposed approach, if we have more than two images even then the same approach is followed but we have to repeat it multiple times by taking one fixed reference image and changing the other image. So in the following subsections we will discuss the approach as though we have only two images.

A. Geometric Alignment: Homography

As we have discussed, if we have images of the scene from different views we can then remove the occluding object and try to fill up the patch with the correct pixels from the other image, so we need to find a geometrical transformation to relate pixels in the two images. A homography transform []is a 3×3 full rank matrix that maps points from one plane to another. It is interesting to note that while if the scene of interest is planar as is the case with many buildings then we can easily map the two views with a homography, also more importantly even if the scene is not planar but if the camera center has not been changed even then both the scenes are related by a homography as proved in[[8]].

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} h_1 1 & h_1 2 & h_1 3 \\ h_2 1 & h_2 2 & h_2 3 \\ h_3 1 & h_3 2 & h_3 3 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} x_1 & y_1 & z_1 \end{pmatrix} - pixel \quad coordinates \quad in \quad 1st \quad image$$

$$\begin{pmatrix} x_2 & y_2 & 1 \end{pmatrix} - pixel \quad coordinates \quad in \quad 2nd \quad image$$

Now if the camera center does not exactly remain the same but moves slightly, then as long as this movement is reasonably small compared to depths of objects in the scene even then we make the assumption that there will still be a reasonably good homography mapping existing between the scenes even but we will have to accomodate a small translation to account for the motion of the camera as well. This kind of a scenario occurs mostly in the ubiquitous case of hand-held mobile cameras. However if the motion of the camera is too much then there will be unpleasant warping induced in the scene so care must be take that the camera center is not shifted by much. If the scene is planar we can go even further in that we can take images of the same scene by some other camera (of course from different view points) as well, this means that we can use the vast amount of images present in the internet as well so this would be a very convenient case for us. Once we have a set of images that satisfy our above criteria then we find the point correspondences between them and then estimate the homography matrix H [[8]]. Once H is estimated between two of the given images we then create a mask such that it will remove the occluding object from the first image (i.e masking it with black), and then using H we align the pixels in the other image. Now mask the first image and blend it with second image accordingly.

B. Blending: Laplacian

We use Laplacian Blending to combine both the aligned images and a manually defined mask that masks the occluding object, for this first the gaussian pyramids are created for both the images and from these we obtain their laplacian pyramids as the difference of gaussians at each level, the gaussian pyramid of the mask is computed and from these the laplacian pyramid for the blended image is calculated for each level i as :

$L_{result,i}=G_{mask,i} \times L_{image1,i}+1-G_{mask,i} \times L_{image2,i}$

$L_{result,i}$: Laplacian pyramid of the result image at level i.

$L_{image1,i}$: Laplacian pyramid of the image 1 at level i.

$L_{image2,i}$: Laplacian pyramid of the image 2 at level i.

C. Luminance factor

If the two images are taken by different cameras or at different times or in different lighting conditions then the same corresponding regions in both the images will have different intensity values so just transferring the pixels from one image to the other will give us a result that is not visually appealing even after laplacian blending. To counter this we introduce a luminance factor, both the images are converted from RGB colorspace to CIELab colorspace then the average luminance is calculated in both the images and the ratio of these is taken which is the Luminance factor. We multiply L channel of the appropriate image with this factor and then perform laplacian blending.

D. Seam Carving

It often happens that the second image might not be able to completely fill in the patch caused by the object that has been removed this leaves some black patches in the result one way to get rid of these is to add seams accordingly, this however was found to be a very unappealing, due to the addition of seams the transferred pixels get shifted in the vertical or horizontal directions and this causes discontinuities in the previously perfectly aligned edges and other features across the patch boundary. So we instead decrease the energy of the unfilled portion of the patch and remove it using seam carving[[9]]. The minimum energy seams (which will include the unfilled portion of the patch) are removed and we get our final result.

IV. Results

Consider the dataset of two images Fig.1 and Fig.2, Our task is to get a complete image of the room, the person is currently occluding the room in both the images, but as the person was moving and his position has changed the information occluded in one image is available in the other image for both the images as the person was moving and his position has changed. Also it is known that the camera was not moved much while in between capturing the two images so we can find a homography transformation H between the two images. the H matrix is found and the second image is aligned with the first image as shown in Fig.3. Also as the images were captured by the same camera and under similar lighting conditions we need not bother about the luminance factor in this case, in fact it was computed and was found to be almost unity and its inclusion did not change the visual perceptibility so it was not considered for this dataset. Initially the pixels are transfered just using the H matrix to obtain the result as

Fig. 1. Image 1.

Fig. 2. Image 2.

Fig. 3. Geometrically aligned image 2.

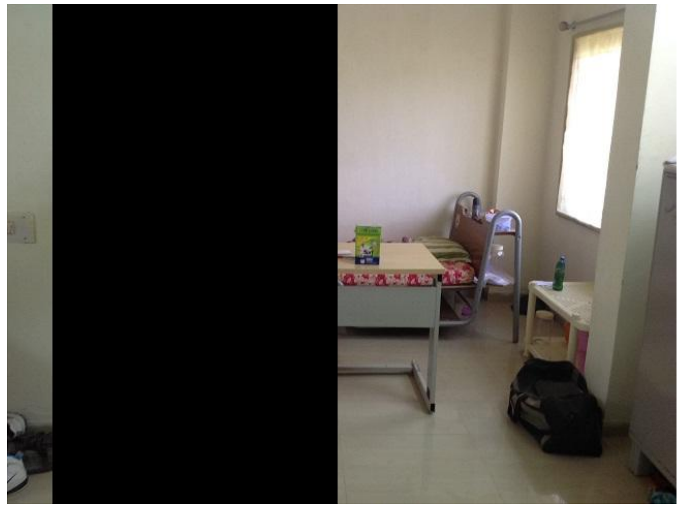

Fig. 4. Image 1 with the patch to be filled.

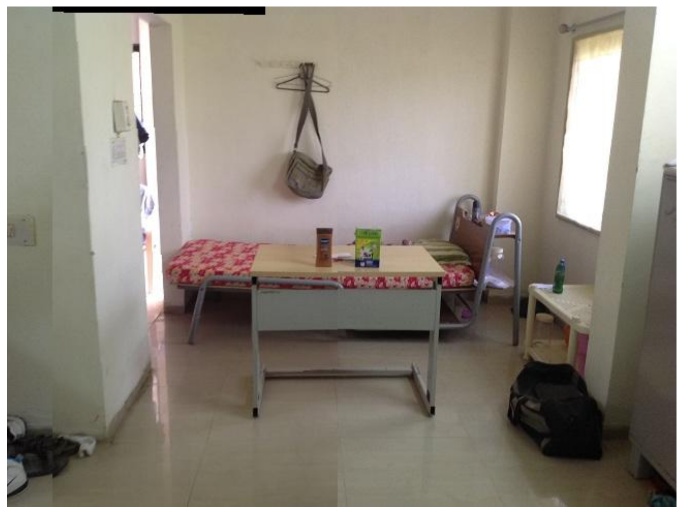

Fig. 5. Result of directly transferring the pixels without Laplacian Blending.

shown in Fig.5. We remove the entire block of the image which has the person in it, here we have manually decided what the patch has to be as shown in Fig.4. As we can see, the patch is replaced by the part of the wall that would actually have been there if not for the occlusion. So we are able to perform the inpainting using homography. But if we observe Fig.5 closely the result is not visually very appealing as the boundary of the patch which was removed are clearly visible and i.e is no smooth transition at the boundary of the patch. So instead of directly transferring the pixels we perform laplacian blending between the geometrically aligned Image 2 and Image 1 with the appropriate mask to remove the patch. We can see the result of this operation in Fig.6. We see that though it is visually much better than Fig.5, Fig.6 still has some parts of the patch unfilled as the image 2 does not have that information. So we perform Seam carving after assigning very low and negative energy to pixels corresponding to unfilled parts of the patch to remove it. Fig.7 is the final result obtained.

Fig. 6. Result after Laplacian Blending.

Fig. 7. Final Result after Seam Carving.

V. CONCLUSION

Thus using this technique we can inpaint regions of any size as long as we have another image which can supply the information regarding the occluded part. We are able to obtain a visually pleasing inpainted image from the two given images after removing the occluding object. Thus using homography proves to be a nice technique for inpainting, other correspondences techniques like NRDC[] were also tried but the results were not satisfactory. One potential improvement could be making the generation of the mask automatic, we can probably achieve this by finding seams along which the intensity variation is very less between the two images as is done in Photomontage[[5]].

REFERENCES

[1] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, “Simultaneous structure and texture image inpainting,” *Image Processing, IEEE Transactions on*, vol. 12, no. 8, pp. 882–889, 2003.

[2] A. A. Efros and T. K. Leung, “Texture synthesis by non-parametric sampling,” in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2. IEEE, 1999, pp. 1033–1038.

[3] A. Criminisi, P. Perez, and K. Toyama, “Object removal by exemplar-based inpainting,” in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2. IEEE, 2003, pp. II–721.

[4] J. Sun, L. Yuan, J. Jia, and H.-Y. Shum, “Image completion with structure propagation,” in *ACM Transactions on Graphics (ToG)*, vol. 24, no. 3. ACM, 2005, pp. 861–868.

[5] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, “Interactive digital photomontage,” *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 294–302, 2004.

[6] C. Rasmussen and T. Korah, “Spatiotemporal inpainting for recovering texture maps of partially occluded building facades,” in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 3. IEEE, 2005, pp. III–125.

[7] M. Wilczkowiak, G. J. Brostow, B. Tordoff, and R. Cipolla, “Hole filling through photomontage,” in *BMVC 2005-Proceedings of the British Machine Vision Conference 2005*. British Machine Vision Association, BMVA, 2005.

[8] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[9] S. Avidan and A. Shamir, “Seam carving for content-aware image resizing,” in *ACM Transactions on graphics (TOG)*, vol. 26, no. 3. ACM, 2007, p. 10.