Appendix B
Adaptive Template Image Matching

One of the primary goals of the UVIS system is to merge together (or mosaic) many images acquired from different locations under a vehicle. The nature of the images to be merged is critically dependent on the details of camera placement and spacing, which also impacts system design and cost. Image matching, the process of identifying like components and their relative displacements in image pairs, provides the foundation for all higher level image processing functions that create image mosaics. The matching task becomes increasingly difficult as the spacing of cameras (relative to their distance from the object to be viewed) increases. As cameras are placed farther apart, the resulting images increasingly differ in appearance due to viewing distortion and even occlusion. The image differences mean that some image components can not be accurately matched with corresponding components in the other image, which complicates the matching process.

However, system design and cost considerations dictate that cameras be placed far apart, or, equivalently, that standoff distances be small, which increases demands on the matching task. Much of the mosaicing work done in the Computer Vision community is based on long range viewing scenarios (such as aerial imagery) where successive images are taken from similar viewpoints, unlike the UVIS system. Therefore, many of the standard matching techniques are not adequate for UVIS without modification. As a result of these considerations, we identified matching as one of the key components of UVIS early in the development process and, consequently, we have been developing enhanced matching techniques that address the challenges of the UVIS requirements.

To appreciate the matching challenges presented by UVIS, consider two successive views of a vehicle taken from locations that are relatively far apart (with respect to offset from the vehicle). Objects will be shifted from one image to the next by an amount that is dependent on their depths from the cameras. This can often result in an arbitrarily large change in background for a given object (i.e. - in one image a given object may have a very dark background and in the next image it could have a bright background since the background components may shift quite differently from the object due to different depths within the scene). The resulting change in image patterns is highly problematic for standard matching techniques. While details vary significantly, most image matching techniques are based on correlation , e.g. moving a small sample from one image over the other image to find the optimum local match. The problem is that the image sample used is fixed (usually a rectangular window) and when it is compared to the other image, there may be a strong match for part of the pattern but another part may not match at all due to the shift in background described above. The result can be a significantly reduced match signal for the correct shift values which degrades the matching output data (see Figure B.1).

We are addressing this problem by developing an alternative matching strategy. Instead of using a fixed window to compare one image region to another, we find the largest pattern that produces a good match. The remaining unmatched area within the image is then matched separately and the process is repeated until no more matches can be found. This naturally decomposes the template into subregions that yield optimal matches and can follow arbitrary boundaries so image components can be separated and treated independently. Another benefit of this approach is that disparity (shift) results can change abruptly so that sharp edges in both 2D and 3D can be recovered without smoothing degradation.

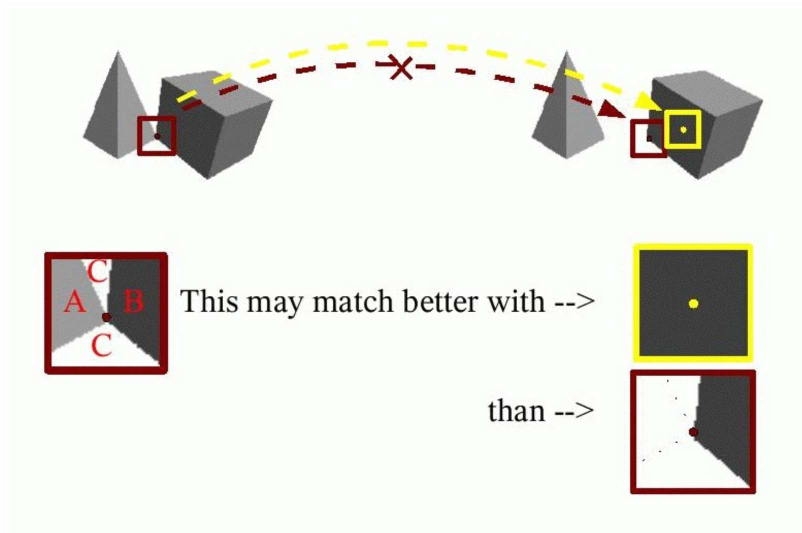


Figure B.1 Changes within Fixed Template Reduce Match Signal. Since match quality in fixed template matching is based on similarity over the entire template region, the match signal is reduced if the template region changes.

While developing this approach, we found that using any fixed window to define an initial match region to be decomposed would introduce arbitrary artifacts, depending on such factors as how window boundaries aligned with image objects. We were able to remove this limitation by treating the entire image as a whole, which is equivalent to using the entire image as the match template window. Our algorithm for matching two images currently consists of holding one image fixed and globally shifting the other image by integer amounts in X and Y over the entire range of possible values defined by imaging geometry. The shifted image is compared with the reference (fixed) image and all matched pixel locations are found. Connected pixels are grouped into regions and the size of each region is determined. The disparity values (shifts in X and Y at every pixel) are built up by placing at every pixel location the value of a given shift in X and Y if the size of the region the pixel belongs to is larger than the previous largest region associated with that pixel.

This process yields the largest image components that match with their corresponding components in the other image and finds the natural boundaries between these components. This eliminates the problem of shifted backgrounds since each image component is naturally decomposed along boundaries that yield the optimum match in terms of region size. The effective match template naturally decomposes along boundaries that optimize matching, in contrast to traditional matching approaches that are restricted to using fixed window templates with the attendant matching data degradation.

The basis for matching in our adaptive template strategy is maximizing matched region size, while conventional fixed template approaches are based on maximizing the total match over the range of the fixed template without regard for contents in the template. With this approach, discontinuous disparity of arbitrary size can readily be recovered without smoothing edges (see Figure B.2).

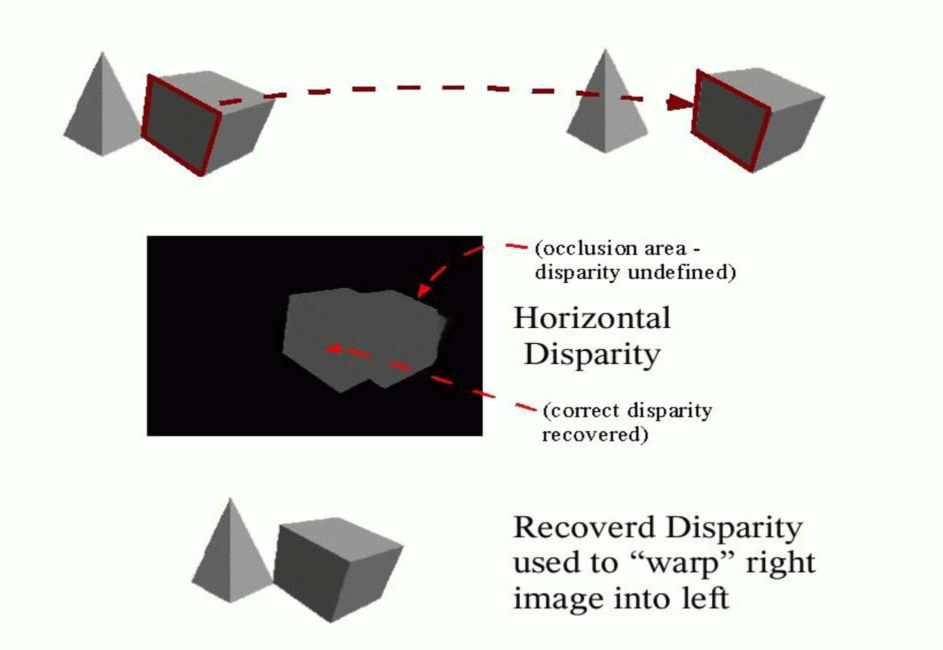


Figure B.2 Adaptive Template Matching Left and right images are shown on top with one of the adaptive templates highlighted. The recovered disparity is shown in the middle image – note the lack of smoothing across edges. The region to the right results from an occlusion area, which has, necessarily, undefined disparity. This is not necessarily a problem when disparity is used to register one image with another to achieve similar appearance, as was done for the bottom image.

It appears that our adaptive template approach correctly recovers disparity and, therefore, depth information as well, when matching patterns are quite similar. Further, it can deal gracefully with distorted image regions by effectively morphing one image region into another. While the associated depth information may not be meaningful in highly distorted regions, image appearances can be effectively matched which still allows for the use of some mosaicing strategies. Another feature of this matching approach is that match region size information is produced and can be used to aid in different processing schemes, as will be demonstrated below.

We have evaluated adaptive template image matching using different image sets and tests. One test verifies that recovered disparity is consistent with manually determined shifts across image pairs. Another test uses the recovered disparity to "morph" one image into the other (as shown in Figure B.3). If the resulting image is similar in appearance to the original, that is one indication that the recovered disparity is correct, in some sense, and can provide useful information. An extension of this evaluation is merging two images matched using the adaptive template process. As one example, we take two real images from a UVIS image set that are beyond the matching capability of the current UVIS system due to the distance between the viewpoints relative to the camera offset distance. The recovered disparity is used to "morph" the second image into the first (see Figure B.4).

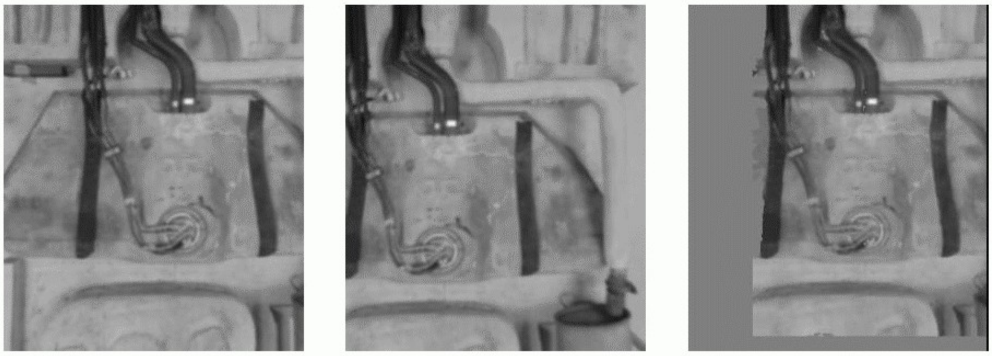
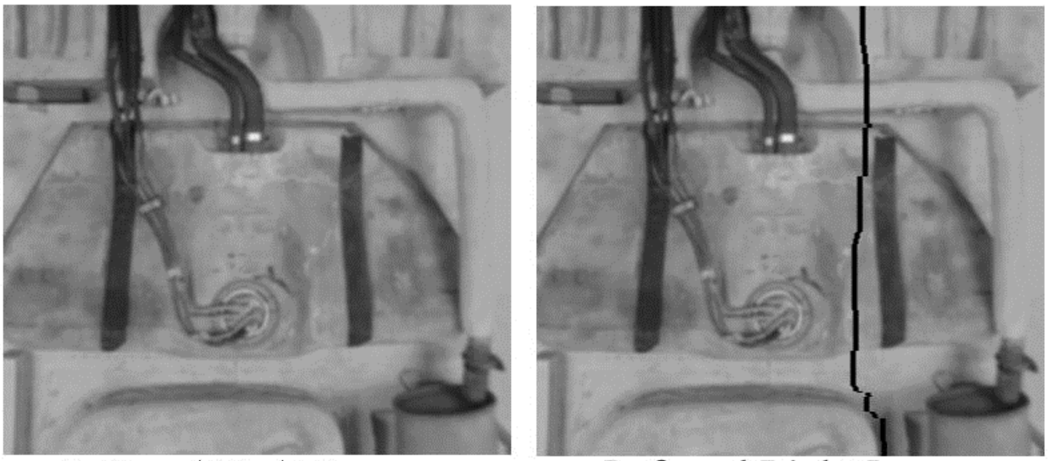


Figure B.3. Adaptive Template Matching applied to UVIS Imagery.
Left and Right images are taken from a real UVIS image set. The images used here are beyond the current matching capabilities of the UVIS system and contain non-overlapping areas. The image on the far right is generated by morphing the right image into the left image using disparity recovered from the adaptive template matching process. Note the curved vertical strip in the registered image which is generated by warping the straighter strip in the right image.

A seam is then adaptively chosen, based on maximizing region size to find minimally disruptive locations. The left image and morphed image are cut along the determined seam and merged at the seam to form a new composite image. Pixels to the left of the seam are taken from the left image and pixels from the right of the seam are taken from the registered image. This is an example of morphing that does not use 3D information, it only merges images along a minimally disruptive seam so geometric integrity is not necessarily preserved. It should be noted that all of the two images are used for merging even though not all areas of the images overlap, which could not be done using a 3D based approach since no 3D information can be determined for image regions that do not overlap.

However, the most relevant evaluation requires integrating the new adaptive matching mechanism into the full UVIS system and determining how (if it all) the results are improved. This involves large and time consuming changes to much of the software in the current UVIS system, so this is ongoing work at this time. If the adaptive matching process can be successfully integrated into the UVIS system, many advantages may be realized. The improved matching process may be able to tolerate larger camera and view distances with respect to camera standoff distance. If so, then improved image quality and/or reduced system footprint could be achieved.

Improved image quality could take the form of fewer artifacts or visible seams, higher fidelity geometric reconstruction, better treatment of edges, and tolerance of more vehicle motion. Alternatively, larger distances between views could result in fewer cameras, therefore, reduced system footprint size and cost.



Left and Right Images Merged (Mosiaced)

Left and Right Images Merged showing Seam

Figure B.4. Simple Mosaicing using Adaptive Template Matching**.** Left image is the result of merging the images in Figure B.3 using adaptive template matching. Even non-overlapping areas are treated. Note the absence of any detectable seam. The image on the right shows the seam used for merging which is found by adaptively seeking large regions extracted by the matching process.