FAST TEMPLATE MATCHING OF REPETITIVE OBJECTS IN STEREOSCOPY

Youval Nehmadi[1], Orly Kalantyrsky[2] and Hugo Guterman[3]
[1] Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[2] Department of Computer Science, Tel Aviv-Yaffo Academic College, Tel Aviv, Israel
[3] Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Keywords: Image Processing, Image Registration, Stereo Vision.

Abstract: One of the challenges of stereovision is to process images with repetitive objects. In order to calculate the distance to an object, matching of the corresponding points between two images must be done. When repetitive objects exist, matching is not straightforward. Many known stereo methods rely on a uniqueness constraint. A uniqueness constraint assumes that only one correct match exists between stereo images. Some algorithms ignore repetitive objects and omit them in the depth map. We present a method that does not employ a uniqueness constraint, but rather determines whether an object is repetitive and then solves the matching problem by finding a unique object that is in close proximity to the object.

1 INTRODUCTION

Image registration (Zitova, 2003) is required in many applications including remote sensing, sensor fusion, stereo vision, panoramic imaging, noise reduction, hyper resolution, 3D imaging. Basically, image registration can be defined as the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. Efficient implementation of the overlaying technique of the two images is especially important for stereo where even small registration errors might greatly affect the construction of the 3D model. Due to its relevance, the topic of image registration and object matching has been widely studied and a variety of approaches had been proposed (Zitova and Flusser (2003), Cyganek and Siebert (2009), Mühlmann, Maier, Hesser and Männer (2002), Shechtman and Irani (2007), Scharstein and Szeliski (2002)). Object based matching methods are widely used in stereovision. Matching of the objects in two stereo images is necessary in order to obtain 3D information on the object. Several of the proposed approaches employ cross-correlation to perform image registration, however this is computationally intensive. Different real-time solutions of the correlation-based registration have been implemented on a variety of hardware.

Generally, registration methods assume two main constraints:

1. The epipolar geometry constraint according to which the corresponding points lay on the epipolar lines of two images.
2. The uniqueness constraint according to which the objects within the image are unique.

While the epipolar constraint can be applied on a calibrated stereo set, the uniqueness constraint presents serious limitations, especially when the information is attained with a set of moving cameras. However, in real scenarios there are many cases where an object inside a region of interest (ROI) does not have a unique appearance, but appears more than once in the search window (Figure 1). In these cases the registration algorithms fail to provide accurate results.

In order to estimate the distance to an object using stereo vision, the object needs to be identified in both stereo images. When a repetitive object exists in one image, it might have several matching objects on the other image. As a result, a wrong object might be selected and the 3D result will be deformed. In order to avoid this deformation we need to recognize repetitive objects and to take them into consideration when performing the matching. In most cases, a correlation algorithm is used to perform image registration and to identify the same

Nehmadi Y., Kalantyrsky O. and Guterman H. (2012).
FAST TEMPLATE MATCHING OF REPETITIVE OBJECTS IN STEREOSCOPY.
In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*, pages 198-205
DOI: 10.5220/0003778501980205

object at corresponding points in the two images of a stereo pair.

Such algorithms are known to fail when:

- there are repetitive objects
- the area has only a little texture
- disparities vary rapidly within the correlation window
- an occlusion exists
- the image does not comply to the ordering constraint (Gong and Yang, 2003).

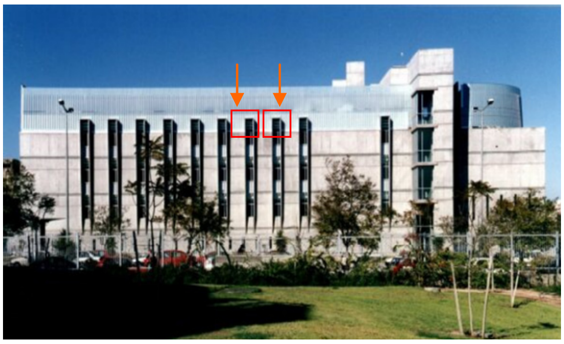


Figure 1: An example of repetitive objects: the windows in the building are repetitive (see red arrows).

Over the years several attempts have been made to overcome these problems (Okutomi and Kanade (1993), Szeliski and Scharstein (2002)). In many cases, the algorithms ignore problematic locations such as repetitive objects or occlusions in order to avoid significant depth errors. However, removing those locations from the calculations is problematic since the distance to these objects is not calculated and is missing in the results.

An example of this approach has been presented by Fua (1993) who uses a consistency criterion to reject invalid matches. The matching is performed twice for each template/pixel. The first time, the template is taken from the first image and matched to the second. The second time, the template is taken from the second image and matched to the first. Only when both matchings result in the same location is the matching considered valid. Otherwise the templates/pixels are rejected. This method rejects repetitive objects and the distance to those objects is not calculated. The advantage in our approach is that instead of rejecting the repetitive objects we find those objects and remove the repetition by adding a location that stops the repetition.

Szeliski and Scharstein (2002) presented an algorithm for stereo matching that addresses two factors - the uniqueness constraint and the stereo occlusions. The algorithm uses the symmetric matching of Fua (1993) to detect ambiguous matching of repetitive objects. It resolves this ambiguity using adaptive window approach that enlarges the template size to include non-repetitive objects (Kanade and Okutomi, 1994). In general, the template should be large enough to include enough texture for correlation matching. On the other hand, it should be small enough to avoid unwanted smoothing and the effects of projection distortion. The probability of mismatching decreases as the size of the template increases. Too small a template will result in poor disparity estimation, since the signal-to-noise ratio is low due to the lack of texture. However, when the template is too large it leads to loss of accuracy due to disparity changes within the template. This causes different projection distortions in both images. In addition, a large window contributes to additional noise from regions without texture (Kanade and Okutomi, 1994). In these cases, the position of the maximum correlation may not represent accurately the correct matching. Kanade and Okutomi (1994) suggested a method for adaptive window size selection. This approach increases the template size iteratively and calculates the uncertainty of matching. The template size increases as long as the uncertainty of matching decreases. The method presented in our paper finds the regions that need to be added to the original template directly without any iterations. Additionally, instead of enlarging the whole template size we add to the template only one region that resolves the matching uncertainty.

In this paper, a new method for dealing with repetitive objects in stereo images is proposed. The proposed method creates a composed template based on multiple small templates that contain relevant information and removes regions that might yield bad results such as regions without texture or regions with large disparity changes. An instance of the repetitive object in combination with the object that breaks the repetition creates a unique composed template. The method is computationally less intensive than most other approaches.

2 METHOD AND IMPLEMENTATION

Feature based stereo techniques match templates from the left image to those in the right. Templates were selected in regions with high intensity variations (edges, corners, etc.). A flow chart of the algorithm is shown in Figure 3. The main steps of the algorithm are described below:

1. Correlation of the template from the left image with the right image.
2. Check how many valid peaks exist in the correlation map. Three options exist:
   i. No peak results in matching. The template location should be omitted from the 3D map. Go to Step 1 for next template.
   ii. One peak identifies a unique matching of the template. Go to Step 1 for next template.
   iii. More than one peak is detected. The template is labeled as “suspected to be repetitive”. In this case the algorithm should continue to Step 3.
3. Verify the repetitiveness of the template on the left image. This part of the algorithm is described in details in section 2.1 below. If the template is confirmed to be repetitive, the algorithm should continue to Step 4, otherwise the template is disqualified and the algorithm continues to Step 1 for next template.
4. Composing the unique template: An additional template that breaks the repetition is added to the original template (see section 0 for details). This composed template is used for correction of the matching in the next step.
5. Correlation of the composed template: the composed template contains the original template and the unique template (found in Step 4). The composed template is used to obtain the matching location as presented in section 0 below.
6. Go to Step 1 for next template.

Figure 3 presents an example of composed template correlation matching. The repetitive template is marked green on the left image. This template was matched using normalized cross-correlation to the right image. Two matches were found on the right image. The first match is marked green and the second match is marked blue on the right image. The algorithm added an additional template that together with the original repetitive template composes unique template. The purpose of an additional template is to break the repetition and to select the correct match among the repetitive matches. The additional template is marked red on the left image.

The correlation of the composed template corrected the matching of the repetitive template. The selected match is marked yellow (same location as the second match marked blue).

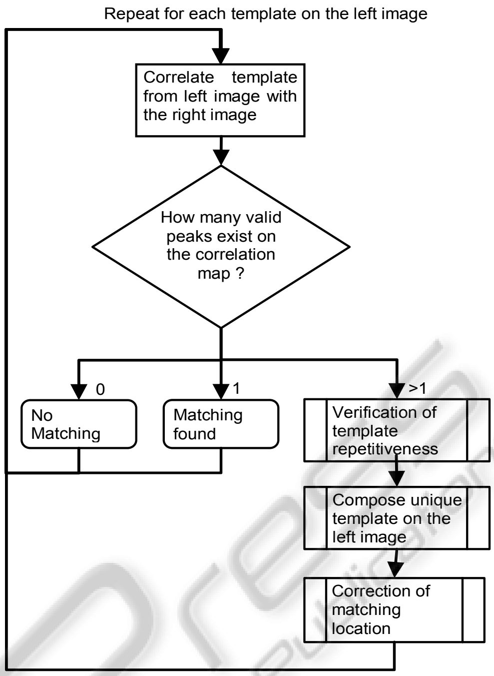


Figure 2: Flow chart.

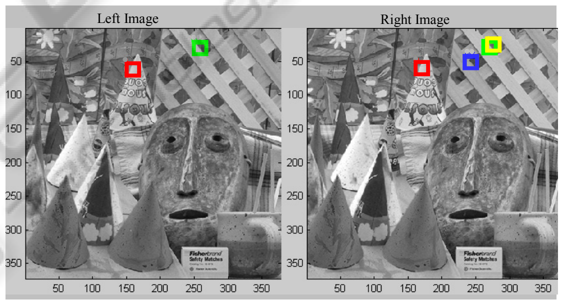


Figure 3: Template location on the left and right images.

2.1 Verification of Template Repetitiveness

In order to find the location of the template from the left image on the right image, normalized cross-correlation is performed. The peaks in the correlation map represent matching. When this template is repetitive there is more than one valid peak in the correlation map. The algorithm checks this by comparing the second maximum value to the first maximum value. If the values are close (e.g. their ratio is bigger than 0.8), the algorithm verifies the repetitiveness of the template on the left image. This time the normalized cross-correlation of the template is performed on the left image. The maximum value in the correlation map identifies the original location of the template. In order to verify template repetitiveness, the algorithm compares the second maximum value to the first maximum value of the correlation map. If the ratio is bigger than predefined threshold (0.7), repetitiveness verification succeeded and the algorithm to reduce repetitiveness is activated as described in section 0.

An example is given in Figure 4 . The location of the first maximum is marked green and the location of the second maximum is marked blue in the first image (Figure 4 (a)). The peaks are marked on the correlation map at the bottom.

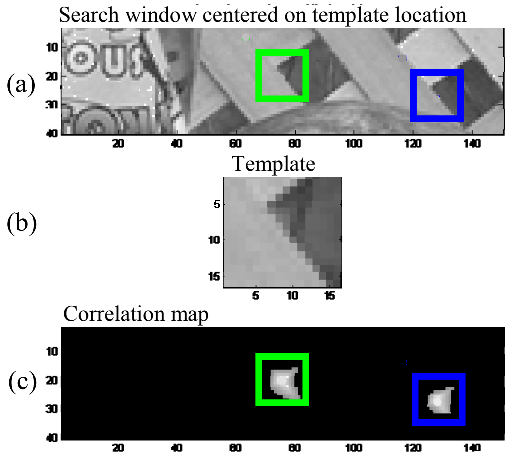


Figure 4: (a) Search window with repetitive template. (b) The template. (c) Correlation map of the template and search window.

2.2 Selection of a Unique Template

A "composed unique template" is composed of the original repetitive template and an additional unique template. This section describes how to find such an additional unique template. The selection of the additional unique template is performed on the original left image.

In order to identify an additional template that would break the repetition, two image fragments have to be clipped from the left image. These image fragments are centered on the first and the second repetition locations of the template, and subtracted one from the other. High values in the result of image subtraction represent locations that do not repeat as frequently as the repetitive templates. The pattern that defines the uniqueness should be selected from the subtraction result in the areas with high values.

Figure 5 shows an example of a schematic image, in which *match 1* and *match 2* are locations that result from the first and second peaks of the correlation. The image fragments are cropped and centered on *match 1* and *match 2* locations. Image fragment 1 contains an object that breaks the repetition. The image fragments are subtracted. The object that breaks the repetition contributes high values to the subtraction result.

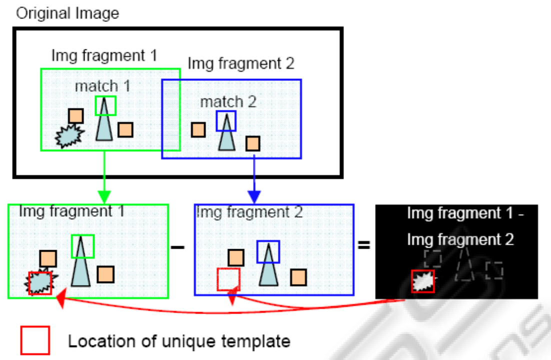


Figure 5: Finding a unique template to avoid repetitions.

An example of a real image is shown in Figure 6. Two image fragments are clipped from the original left image and centered on the template repetition locations – on the first and second peaks. The image fragments are shown in Figure 6 (a)-(b). The bottom image represents the subtraction of these two image fragments. The high values (bright points) on the subtraction are locations that do not repeat with the same frequency as the repetitive template.

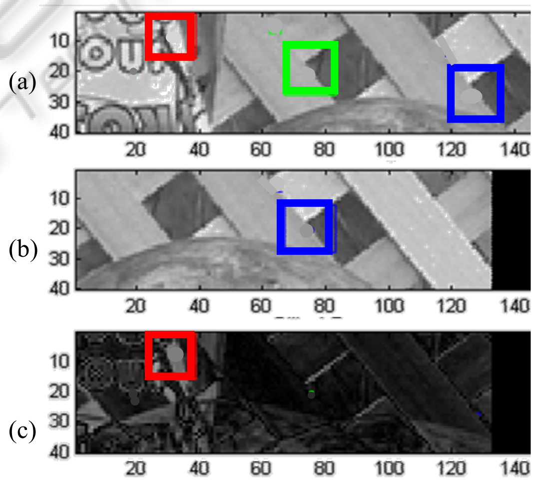


Figure 6: Subtraction of image fragments with repetitive template. (a) Image fragment centered on first maximum location is marked in green. (b) Image fragment centered on second maximum location is marked in blue. (c) The subtraction result. The red mark represents maximum value in the subtraction result.

Two additional conditions are important for template selection:

1. For better correlation results, an additional template should be selected in the area that

contains patterns (edges/corners).

2. To minimize distortion caused by the different perspectives of the stereo vision, the unique template should be selected close to the original template location (first maximum).

2.3 Correlation of the Composed Template

In the previous section we described how to compose a unique template on the left image. An additional template that breaks the repetition was added to the original repetitive template. This section describes how to correlate the composed unique template with the right image to obtain the matching of the repetitive object.

An additional template was selected in the neighborhood of the original repetitive template. We assumed that stereo distortion did not have a significant effect on the distance between these two objects within the stereo images. This means that the distance in pixels between the original repetitive template and the unique template is similar in both images.

The matching of the templates is performed by normalized cross-correlation, which selects search windows on the right image.

Two search windows for matching both templates are clipped from the left image. The search windows are centered on the coordinates of the templates, according to their original location on the left image. An example is shown in

Figure 7, where the selected repetitive template coordinates within the left image (x_1,y_1) are marked in green on the left image. The search window on the right image is centered (x_1,y_1), where it appears as a blue (yellow) rectangle on the right image. A search window for the unique pattern is similarly selected. In the figure, the unique pattern coordinates (x_2,y_2) on the left image are marked red. The search window on the right image is selected with the center on (x_2,y_2) on the right image. It appears as a pink rectangle.

The matching of the templates and their search windows is performed by normalized cross-correlation as defined below.

$$C_{xy} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}, \quad \bar{x} = \frac{x}{N}, \; \bar{y} = \frac{y}{N} \quad (1)$$

The normalized correlation result is a map with values between 0 and 1.

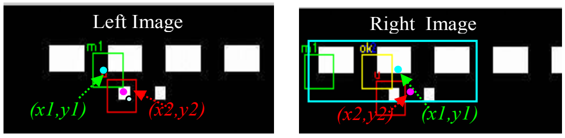


Figure 7: Selection of search windows on the right image.

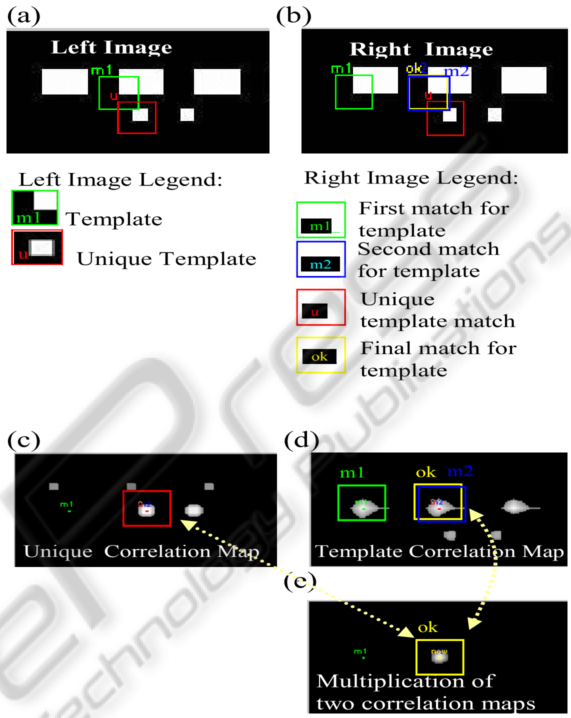


Figure 8: Combining templates by element-by-element multiplication of two correlation maps. (a) On the left image the template is marked green and the unique template is marked red. (b) Right image, where first matching peak marked green, second peak – blue, third peak – pink. The unique template is marked in red and the final repetitive match is selected in the location marked in yellow. (c) Correlation map of the unique template. (d) Correlation map of the template. (e) Element-by-element multiplication between two correlation maps (c) and (d).

In order to select the correct matching of the repetitive pattern, the element-by-element multiplication of the correlation map in the repetitive template and its search window is calculated. The multiplication of both correlations removes redundant maximums (Figure 8(e)). This process enables us to correct the template location. Element-by-element multiplication of two normalized cross-correlation maps results also in a map with values between 0 and 1. This result is close to 1 if two combined templates were perfectly matched and their stereo displacement was equal, but would be close to 0 if the templates do not match (see Figure 8). The repetitive template is marked in green and the unique template is marked in red on the left image (Figure 8(a)). The correlation between the repetitive template and the right image (Figure 8(b)) results in three peaks, which are shown in Figure 8(d). The correlation between the unique template and the right image results in two peaks, which are shown in Figure 8(c). Element-by-element multiplication of the two correlation maps (Figure 8(e)) results in one peak only, which identifies the registration between the two images.

The combining template algorithm is calculated as:

$$C_{(t_1 \oplus t_2)y} \cong C_{t_1 y} \bullet C_{t_2 y} \qquad (2)$$

where t_1, t_2 are two templates, $C_{t_1 y}$ $C_{t_2 y}$ are two correlation maps of template t_1 with image y and template t_2 and image y respectively. $C_{t_1 y} \bullet C_{t_2 y}$ is element by element multiplication of $C_{t_1 y}$ and $C_{t_2 y}$. The $C_{(t_1 \oplus t_2)y}$ represents correlation between the template combined from t_1 and t_2 with the image y.

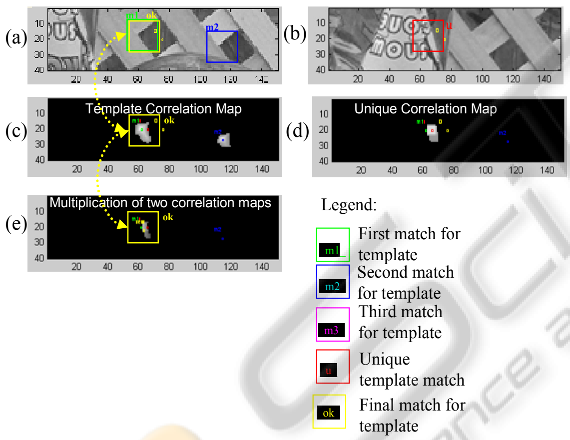


Figure 9: Correction of template location. (a) Image fragment centered on first peak location. (b) Image fragment centered on unique template location. (c) Correlation map template. (d) Correlation map for the unique template. (e) Element-by-element multiplication of two correlation maps (c) and (d).

An example of a real image is shown in Figure 9: (a) - the search window for the repetitive template, (b) - the search window for the unique pattern, (c) - the correlation map of the repetitive template and the search window, (d) - the correlation map of the unique template and its search window. The element-by-element multiplication of both correlation maps is shown in Figure 9(e). The highest value location on the multiplication identifies the peak that represents the matched template location.

3 EXPERIMENTAL RESULTS

The effectiveness of the proposed method was tested. The accuracy of the matching results and the computational complexity were evaluated.

3.1 Algorithm Accuracy

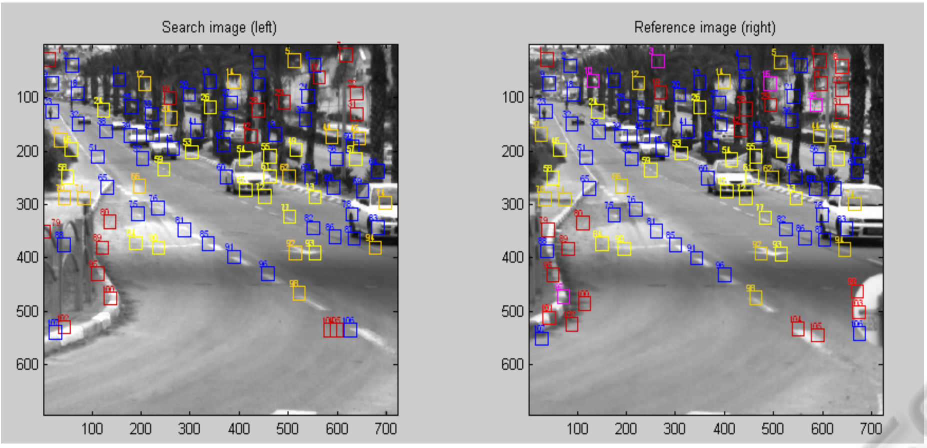
The algorithm identifies templates on the left image and performs the matching on the right image. The method for matching repetitive templates described in section 2 was applied on the templates. Every matching was reviewed manually and acknowledged as correct matching or failure. Table 1 shows the number of templates that were selected on the left image and matched to the right image. The templates are divided into two categories: repetitive and non-repetitive. Table 1 represents the results of the experiment of a real stereo pair. The table represents the results for the matching performed with template size of 5x5 pixels on a real image with the size of 700x700 pixels.

Table 1: Results for stereo matching on real image.

	Template Count	Success Rate
Non-repetitive templates	48	92%
Repetitive templates	33	94%
Total templates	81	93%

3.2 Algorithm Complexity

Calculation time of the template matching is a major limitation in real time implementation. The computation time is dependent in a square ratio with the template size. Using two small templates instead of one large template can significantly reduce the calculation time. An example of the usage of two small templates instead of one large template is shown in Figure 11. The repetitive template is marked green. The additional template selected by the method is marked red. The small templates have the size of 20x20 pixels. Known methods that do not deal with repetitive images would have to select a larger template size in order to include regions that are not repetitive. The large template in this example is marked in pink. The computational ratio in this example is 1/45. In many cases of typical urban scenes we observed a ratio of 1/40.

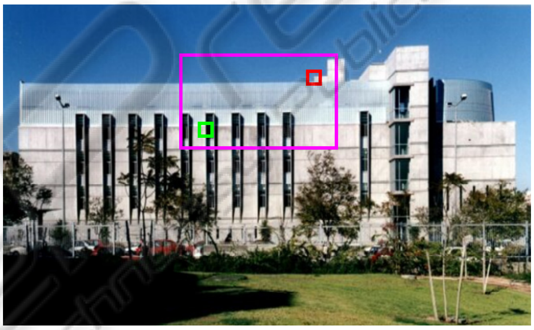


Legend: ☐ - non repetitive templates; ☐ ☐ - repetitive templates

Figure 10: Stereo matching results on real stereo images.

Kanade and Okutomi (1994) described a method that increases the template size sufficiently to include high local intensity variations but with low disparity. The method can be used to enlarge the template in order to include objects that break the repetitiveness. The disadvantage of Kanade's method is the increasing complexity due to the large template size. The template size has a direct effect on the matching complexity. The complexity of the normalized cross correlation is $O(m \times n \times M \times N)$, where the template size is $m \times n$ and the search window size is $M \times N$.

The method presented in this section can be used for improving Kanade's methodology and reduce its complexity. Instead of enlarging the template we use only a subset of the large template by selecting an additional small template which contains a non-repetitive area that breaks the repetition of the original repetitive template. The complexity of template correlation is $N \times M$, where N is the template size and M is the search window. Kanade's method results in a complexity of $P \times N \times M$, where P is the number of iterations required. In the methodology presented in this article we use only two small templates, hence the complexity is $2Ns \times M$, where the size of the small template is Ns. Generally the unique template is located at a certain distance from the repetitive template, therefore $Ns << N$ and the complexity of the method presented here is significantly better than the adaptive template size approach of Kanade (1994).



$\text{Ratio} = \frac{1}{45} \cong \frac{2 \times 20 \times 20}{250 \times 150}$ Large template size 250 X 150
Small templates size 20 X 20

Figure 11: Urban scene with repetitive objects.

4 CONCLUSIONS

In this paper a novel method for pattern recognition of repetitive templates has been presented. When applied to stereo imaging the proposed method solves the matching aspects for repetitive templates. Most stereo algorithms either ignore repetitive patterns or fail to identify them. Algorithms that address repetitive templates dynamically enlarge the template size in order to include unique areas. The presented method is based on identifying an additional pattern that in combination with the repetitive pattern creates a unique template.

By using small templates this novel method addresses the problem of computational efficiency. Instead of performing correlation on large templates, this method uses a unique pattern constructed from two small templates. Usage of small templates is more efficient in computational aspects, for example for computing cross-correlation. Normalized cross-correlation matching has a complexity of $N^2 \cdot K^2$ for the search window with size of *NxN* and with template size of *KxK*. Adding an additional template would require $2N^2 \cdot K^2$ computations instead of $N^2 \cdot L^2$, where $L > K$ and $L^2 >> 2K^2$. In addition to the computational advantage, matching of small features results in lower noise. Matching of featureless regions causes noisy results. In the presented method small templates are selected in high density variation areas, hence less featureless regions are reflected in the correlation.

REFERENCES

Brown, M. Z., Burschka, D. and Hager, G. D. (2003). Advances in Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 993-1008.

Cyganek, B. and Siebert, J. P. (2009). *An Introduction to 3D Computer Vision Techniques and Algorithms*, Wiley.

Fua, P. (1993). A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features. *Machine Vision and Applications*, 6, 35-49.

Hirschmüller, H. and Scharstein, D. (2007). Evaluation Of Cost Functions For Stereo Matching. *IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.

Gong, M. and Yang, Y. H. (2003). Fast Stereo Matching Using Reliability-Based Dynamic Programming and Consistency Constraints. *Proceedings of the 9th IEEE International Conference on Computer Vision*, 1, 610-612.

Kanade, T. and Okutomi, M. (1994). A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 920-932.

Mühlmann, K., Maier, D., Hesser, J. and Männer, R. (2002). Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation. *International Journal of Computer Vision*, 47, 79-88.

Okutomi, M. and Kanade, T. (1993). A Multiple-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15, 353-363.

Scharstein, D. and Pal, C. (2007). Learning Conditional Random Fields for Stereo. *IEEE Conference on Computer Vision and Pattern Recognition*.

Scharstein, D. and Szeliski, R. (2002). A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47, 7-42.

Shechtman, E. and Irani, M. (2007). Matching Local Self-Similarities across Images and Videos, *IEEE Conference on Computer Vision and Pattern Recognition*, 511–518.

Szeliski, R. and Scharstein, D. (2002). Symmetric Sub-Pixel Stereo Matching. *Proceedings of the 7th European Conference on Computer Vision – Part II*, 525-540.

Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11), 977-1000.