

# BUILDING DETECTION AND RECONSTRUCTION FROM AERIAL IMAGES

Dong-Min Woo<sup>a</sup> \*, Quoc-Dat Nguyen<sup>a</sup>, Quang-Dung Nguyen Tran<sup>a</sup>, Dong-Chul Park<sup>a</sup>, Young-Kee Jung<sup>b</sup>

<sup>a</sup> Dept. of Information Engineering, Myongji University, Gyeonggi-do, Korea - (dmwoo, datnguyen, qdungtran, parkd)@mju.ac.kr

<sup>b</sup> Dept. of Computer Engineering, Honam University, -Kwangju, Korea - ykjung@honam.ac.kr

Commission III, ThS-7

**KEY WORDS:** Building, Detection, Reconstruction, Aerial, Image, Feature

## ABSTRACT:

This paper presents a new method for building detection and reconstruction from aerial images. In our approach, we extract the useful building location information from the generated disparity map to segment the interested objects and consequently reduce unnecessary line segments extracted in low level feature extraction step. Hypothesis selection is carried out by using undirected graph, in which close cycles represent complete rooftops hypotheses. By using undirected graph, hypothesis selection becomes a simple graph search for close cycles. This significantly improves the performance of the system over the traditional hypothesis selection methods. We test the proposed method with the synthetic images generated from Avenches dataset of Ascona aerial images. The experiment result shows that our method can be efficiently used for the task of building detection and reconstruction from aerial images.

## 1. INTRODUCTION

The building detection and reconstruction from aerial images is one of the challenge tasks in computer vision. It has been used widely in various applications including traditional applications such as cartography and photo-interpretation and recent application including mission planning, urban planning, computer graphics and virtual reality. There are two main problems needed to solve in any buildings detection approach. The interested objects need to be segmented from the background, and the fragmented line segments of the interested objects should be grouped to human-made structures. These tasks are very challenging, because the objects of interest could be partly occluded by the presence of vegetation, shadows, road and other objects. Moreover, lines and corners of object are often fragmented and missed due to the typical failures of low level features extraction.

Early approaches tried to use a single image only (Huertas, 1988; Lin, 1998). Since the inference of 3D information from one image is very difficult and there are still some ambiguities in the detected buildings that can be only resolved by feature matching in multiple images, the application of the single-image approach is very limited. In this context, most of the recent work in this area has focused on the multiple-view analysis (Fischer, 1998; Noronha, 2001; Collins, 1998). Mohan and Nevatia (Mohan, 1989) proposed an approach for detecting and describing buildings in aerial image using perceptual grouping. They demonstrated the usefulness of the structural relationships called collated features which can be explored by perceptual organization in complex image analysis. All reasonable feature groupings are first detected and the candidates are then selected by a constraint satisfaction network. But this approach involves all extracted line segments in the image. Consequently it costs a big computational effort. It also depends on the accurate extraction of line segments.

Some approaches such as Lin (Lin, 1998) and Noronha (Noronha, 2001) used hypothesis and verification paradigm based on perceptual grouping. Hypotheses are generated by a hierarchical perceptual grouping process and verified by the evidence of visible walls and expected shadows. But the system needs to make several decisions in the selection and verification process based on simplicity and intuitive judgments. In monocular analysis, Jaynes (Jaynes, 1994) proposed feature relation graph, in which hypothesis selection takes places as a graph search problem. This approach improved the performance of hypothesis selection step. However, it is limited on rectangular buildings and tends to generate false hypotheses in complexity images.

In this paper, we extract the suspected building regions in the disparity map generated from aerial images and utilize them to get the location of interested objects in the image. The suspected building regions are areas, where pixel values rapidly changes relative to the surround area. This process can reduce the unnecessary line segments from the low level feature extraction result. Also, we employ perceptual grouping to build collated features, which are used to generate rooftop hypotheses. This perceptual grouping process removes the unsatisfied line segments during the grouping process. Corners and line segments in collated features are used to build the undirected feature graph, in which close cycles are detected as rooftops. With the open connectivity rules between nodes in the graph, our approach can effectively detect rectilinear shape buildings.

The remainder of this paper is organized as follow: an overview of the system is shown in Section 2. Section 3 describes the generation of suspected building region. In Section 4, the principle of low level features processing and rooftop hypothesis is introduced. Section 5 presents experimental results on an aerial image data set. Conclusions are given in Section 6.

\* Corresponding author.

## 2. SYSTEM OVERVIEW

Figure 1 shows the main components in our system. The epipolar images are generated from the aerial images by epipolar resampling process. We obtain the disparity map between the epipolar pairs by stereo matching using area-based matching with non-parametric technique. From the disparity map, we generate the DEM as a 3D terrain model. The building location information extracted from disparity map is used to remove the unnecessary line segments extracted in the low level process. After 2D lines are generated, perceptual grouping is applied to the filtered line segments in order to obtain the structural relationship features such as parallel line segment pairs and U-shapes. These can be used to generate rooftop hypotheses. Among the generated hypothesis, the candidate rooftop is selected by searching close cycles in the undirected graph. Finally, we retrieve 3D buildings by using 3D triangulation for each line segment of detected rooftops.

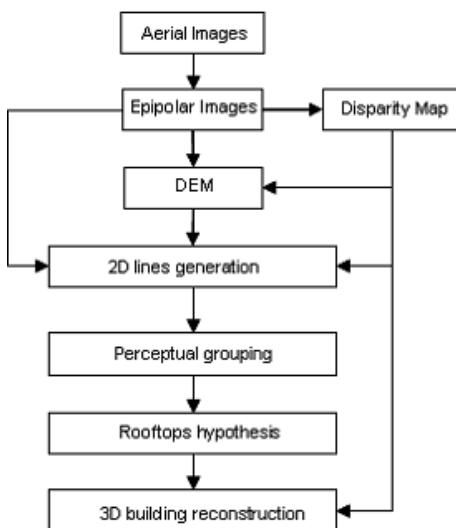


Figure 1. System Overview

## 3. BUILDING REGION EXTRACTION

### 3.1 Stereo Matching

The goal of stereo matching process is to find a match between the pixels in the first (reference) R and second (wrap) W image such that the pixel located at  $(i, j)$  in the R image and a pixel located at  $(i+I(i, j), j+J(i, j))$  in the W image view the same point in object space, i.e.,  $W(i+I(i, j), j+J(i, j)) \rightarrow R(i, j)$ , where  $I(i, j)$  is horizontal disparity map, and  $J(i, j)$  is vertical disparity map. The index  $i$  (column index) is measured along scan lines and the index  $j$  (row index) is measured across scan lines. In this paper, we use epipolar resampled images, and  $J(i, j) = 0$  for all  $i$  and  $j$ . This relation can be reduced to  $W(i+I(i, j), j) \rightarrow R(i, j)$ .

Considering the correspondence problem, there are two popular approaches. The first one is Normalized Cross Correlation (NCC) which is one of area-based matching typical metric, and the second one is non-parametric technique with census transform (Zabih, 1998). We employ the census transform, due to its preservation of the edges and computational simplicity.

To find accurate disparity map, we employed a multi-resolution scheme, referred to as hierarchical, or pyramid processing. For each resolution scheme, the correspondence problem is solved by first computing census transformed image and then using Hamming distance correlation on the transformed image. The census transformation maps the local region surrounding a pixel to a bit string represent which pixels have lesser intensities. For example, in a window surrounding a pixel, if a particular pixel's value is less than the centre pixel, the corresponding position in the bit string will be set to 1, otherwise it is set to 0. After that, two census transformed images will be compared using a similarity metric based on the Hamming distance which is the number of bits that differ in the two correlation window bit string. The Hamming distance (Banks, 1997) is summed over the window:

$$\sum_{(u,v) \in W} Hammi(I_1^c(u, v), I_2^c(x+u, y+v)) \quad (1)$$

where  $I_1^c$ ,  $I_2^c$  represent the census transforms of  $I_1$  and  $I_2$ ,  $W$  is the correlation window.

### 3.2 Suspected Building Region Extraction

It is usually difficult to separate interested objects from 2D line segments collection obtained in low level features extraction. The boundary of interested objects, the buildings, can be partly occluded by vegetation, shadows, and other objects. In rooftop hypothesis process, these fragmented boundaries and the presence of roads, vehicles ... can make false hypotheses including unwanted rooftop and wrong shape rooftop. This causes not only significant computational effort in processing but also wrong final results. To solve this problem, the system should be able to detect line segments that are within or near buildings in the image. Here, we use suspected building regions which extracted from the disparity map. The suspected building regions are areas which pixel values changes in comparison with the surround area. The different of pixel values between suspected building region and surround areas indicates the different of elevation values. It indicates the existing of higher objects such as buildings, trees ... in those regions. In the other words, these regions can give us the information of where the buildings are located.

These regions could be extracted by using a simple height threshold technique. Their boundaries are extracted by convolving the disparity map with a Laplacian-of-Gaussian filter then employing connected component analysis to get zero-crossing pixels' coordinate in the convolution output. We have LoG as an operator or convolution kernel defined as:

$$\begin{aligned}
 LoG(x, y) &= \Delta G_\sigma(x, y) \\
 &= \frac{\partial^2}{\partial x^2} G_\sigma(x, y) + \frac{\partial^2}{\partial y^2} G_\sigma(x, y) \\
 &= -\frac{1}{\pi \sigma^2} \left[ 1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2+y^2}{2\sigma^2}}
 \end{aligned} \quad (2)$$

## 4. 3D ROOFTOP MODEL GENERATION

### 4.1 Low Level Features Extraction

To detect 2D lines from epipolar image, edge detection is carried out first and then 2D lines are formed from edges. We employed Canny edge detector, since it is optimal according to the criteria where edge is defined and comes up with thin edges. To obtain 2D line segment, we use Boldt algorithm (Boldt, 1989) based on token grouping. The method extracts a basic line element, token, in terms of the properties of line A and construct 2D line using grouping process. It is efficient in detecting 2D lines of large structure appeared in urban image.

### 4.2 Grouping and Filtering Processes

Suspected building regions are used to remove line segment that outside or far from interested object boundaries meanwhile the needed line segments are still kept for later processing steps.

Then, we group the closely parallel linear segments since they usually represent a linear structure of objects in image, like the border of a roof or the divider between ground terrain and building, by using a “folding space” between two line segments. If both line segments are inside the folding space, two line segments can be replaced by a single line which orientation is the longer line segment orientation and length is total length of two segments. After this process, each group of the closely overlapping and parallel line segments is represented only by one single line.

Figure 2 showed the typical case of closely parallel linear segments grouping. These linear segments are or nearly parallel lines. So the first condition is the angle between them should be from  $0^{\circ}$  to  $10^{\circ}$ . If two line segments are fragmented lines from one edge, these line segments must be close and should be inside a folding space created by them.

The U shaped structure in Figure 3 is used to detect candidates for rooftop hypothesis generation. Any line segment in a set of parallel lines with aligned end is a U shaped structure candidate which is kept as input for hypothesis generation, otherwise that line segment will be removed.

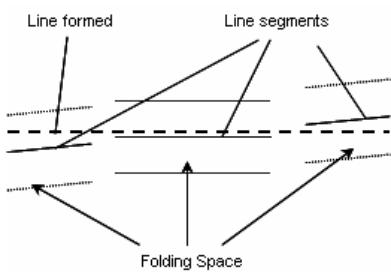


Figure 2. Folding space

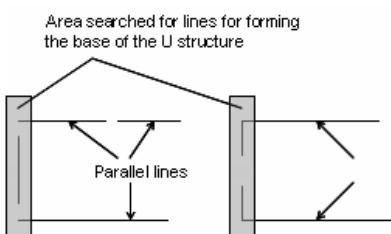


Figure 3 U-structure

### 4.3 Corner Detection

Corner is calculated as intersection of two line segments which their angle is from  $80^{\circ}$  to  $100^{\circ}$  and one of them has nearest distance to another one. We define four types of corner. They are labeled as I, II, III and IV, as shown in Figure 4. Each corner has an attribute to indicate whether it is L-junction or T-junction. This attribute is used to decide whether two different corners have a connection or not. For example, if a corner's label is I and type is L-junction, it connects to any type of corner. However, it prefers connecting to a corner which label is II or IV. If that corner is T-junction, it can only connect to a corner which label is II or IV. This rule is used in hypothesis generation to build collated features.

With the flexible connection between corners, our method is able to detect rectilinear rooftops. Figure 5 show some examples of corner detection, A, B, E, F, G are L-junctions while C, D are T-junctions.

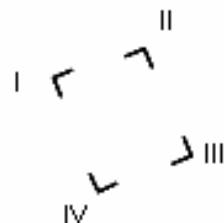


Figure 4. Corner labeling

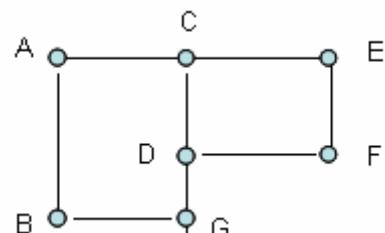


Figure 5. Corner detection

### 4.4 Rooftop Hypothesis Generation

A collated feature is a sequence of perceptually grouped corners and line segments. Here, collated features are constructed from filtered line segments and corners obtained from the filtering and grouping process. That reduces computational effort and false hypotheses.

Hypotheses are formed by alternation of corners and line segments that form collated features. In a collated feature, two corners have connectivity only if they satisfy the corner relation condition and they are the nearest appropriate corner to each other. Beside, every corner connects to only one corner on each its line segment direction. Hypothesis generation is performed by constructing the feature graph. Construction of the graph can be seen as placing corners as nodes and edges between nodes if there is the relation between the corresponding corners in the collated features. When a node is inserted into the graph, the system looks into the remaining nodes whether any node has the relation with the inserted node. If some nodes satisfy the connectivity relation rules, those nodes are inserted into the

graph and the system creates an edge between them. In the example shown in Figure 5, C is T-junction, and it can connect to D, A and E. Meanwhile, A can connect to B, C and E but C is nearer than E towards A on the line segment AE so that A only connects to B and C. Consequently there will be two collated features ACGB and CEF in the Figure 5.

#### 4.5 Rooftop hypothesis selection

The graph is the place to store features and their groupings. Feature as corner is node in the graph and relations between corners are represented with an edge between the corresponding nodes. Closed cycles in the graph represent the rooftop candidates. The hypothesis selection can be seen as a simple graph search problem. The close cycles in the graph are rooftops that we need to detect. Figure 6 show a graph constructed from the example in Figure 5. Corner C and corner D are T-junctions so that there are two nodes in the graph for each corner. Node C1, C2 for corner C and node D1, D2 for corner D. There are two close cycles C1 and C2 as shown in Figure 6.

### 5. EXPERIMENTAL RESULTS

The experimental environment was set up based on Ascona aerial images of the Avenches area. Since this area's 3D model is supplied as a ground truth data, we can evaluate the quantitative accuracy for the 3D rooftop model generated by the proposed method. Two aerial images as shown in Figure 7 are used as a set of stereo images for experiments.

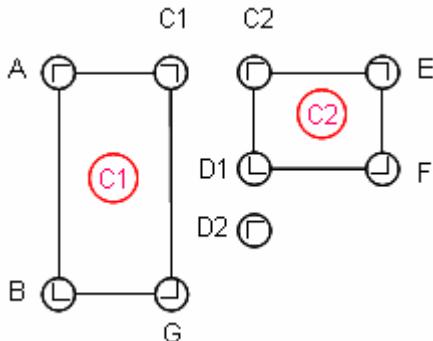


Figure 6. Feature graph



Figure 7. Aerial images used as a set of stereo image

The result of epipolar resampling process is shown in Figure 8. We generate suspected building regions from the map to reduce unnecessary line segments. To find the accurate disparity map, we employed the multi-resolution scheme with four different resolutions, where scaled image sizes equal to original size divided by  $2^n$ ,  $n = (0,1,2,3)$ . The corresponding correlation window sizes are 3x3, 5x5, 7x7 and 9x9, while the census

transform window sizes are 3x3, 5x5, 7x7, and 9x9. The final disparity map is shown in Figure 9. Figure 10 shows generated DEM image and ground truth image.



Figure 8. Example of epipolar images

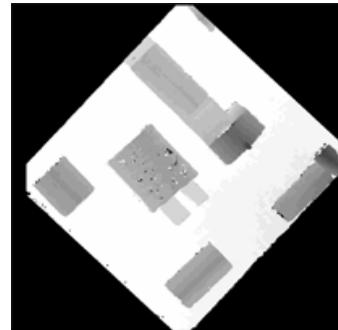


Figure 9. Example of disparity map

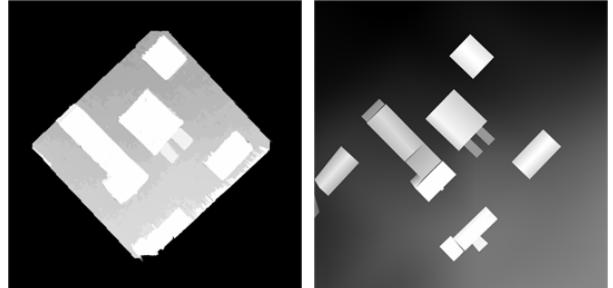


Figure 10. Generated DEM image and ground truth image

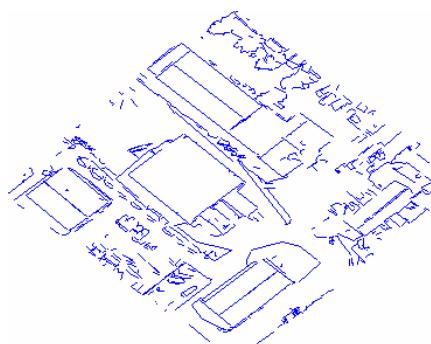


Figure 11. Example of low level extraction result

Figure 11 shows the line segments obtained from low level feature extraction process. The number of extracted line segment is about 1425. To remove unnecessary line segments, we use the suspected building regions extracted from the disparity map as shown in Figure 12.



Figure 12. Example of suspected building regions

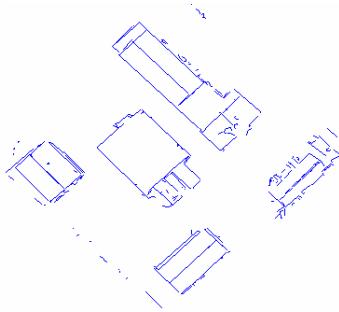


Figure 13. Example of filtered line segments

The result of unnecessary line segments removing is shown in Figure 13. Remaining line segments are about 405. Now, perceptual filtering and grouping process is employed to obtain line segments which can be part of any U-structure group. The close parallel line segments which are inside their folding space of each other will be grouped into one representation line. The line segments which are part of a collection of line segments forming U-structure will be used to generate hypotheses in the next step. Figure 14 shows the line segments forming U-structures in a collection of line segments. The colors indicate which U-structure group that the line segment belongs to.

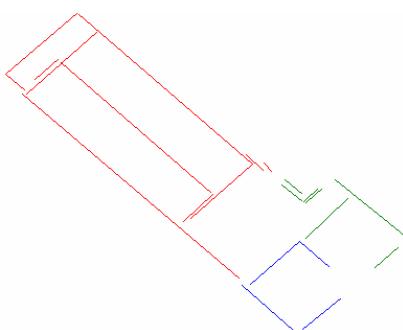


Figure 14. Example of U-structures

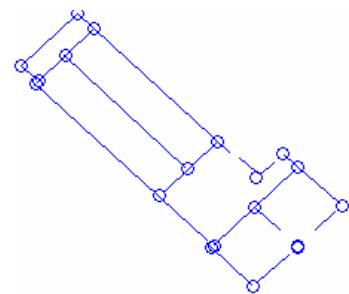


Figure 15. Example of corner detection

The corners are calculated form the intersection of the line segments which satisfy two conditions: their angle is from  $85^0$  to  $95^0$  and one of them has nearest distance to another one. Figure 15 shows extracted corners from the line segments collection.

Using the obtained corners and line segments from the previous steps, we can build the collated features. In order to have a link between each other, two corners must satisfy the connecting relation of corner type and the required condition of their distance. Another important rule that help to define the corner connectivity is on each line segment of a corner, there is only one corner has connection with it. Figure 16 shows the collated features obtained from the line segments collection.

The collated features are used to construct graph by placing a corner as a nodes and line segment two corners as edge between two nodes if there is the relation between the corresponding corners in the collated features. Closed cycles in the graph represent the possible rooftops. Hypothesis selection becomes the searching of close cycles in the graph. Figure 17 shows the close cycles selected from the line segments collection.

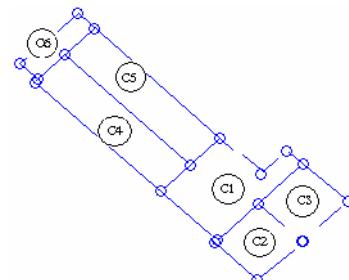


Figure 16. Example of collated features

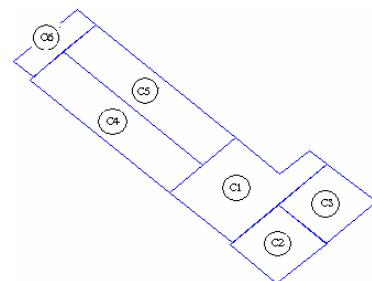


Figure 17. Example of close cycles

Figure 18 shows rooftop detection results of the entire area. There is a building located near the border line of the epipolar image that the system can not detect correctly due to missing line segments in low level extraction step. The remaining buildings are accurately extracted. From the detected 3D rooftop, we generate a 3D rendering view, as shown in Figure 19. The 3D view indicates that the detected 3D rooftop accurately reflects ground truth 3D model.

## 6. CONCLUSIONS

A new approach to detect and reconstruct buildings using perceptual organization from two aerial images has been suggested. Low level feature extraction is performed in the epipolar images, which makes it possible to reduce the search effort in matching process. The proposed suspected building regions can be utilized to remove the unnecessary line segments, prior to the generation of rooftop hypotheses. This region can reduce computational complexity and false hypotheses. Using undirected feature graph, the selection of rooftop hypotheses becomes a simple graph searching for close cycles. Experimental result shows that the proposed method can be very effectively utilized for the rectilinear structures of urban area.

## ACKNOWLEDGEMENTS

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MOST)(Grant No.: R01-2007-000-20330-0).



Figure 18. Example of detected rooftops

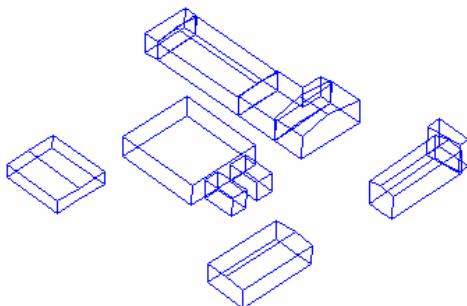


Figure 19. 3D view of building rooftop

## REFERENCES

- Banks, J., Bennamoun, M., Corke, P., 1997. Non-parametric technique for fast and robust stereo matching. *Proc. of IEEE TENCON*, pp. 365-368.
- Boldt, M., Weiss, R., Riseman, E., 1989. Token-based extraction of straight lines. *IEEE Transaction on Systems Man Cybernetics*, 19(6), pp. 1581-1594.
- Collins, R., Jaynes, C., Cheng, Y.Q., Wang, X., Stolle, F., Riseman, E., Hanson, A., 1998. The Ascender system: automated site modeling from multiple aerial images. *Computer Vision and Image Understanding*, 72(2), pp. 143-162.
- Fischer, A., Kolbe, T., Lang, F., Cremers, A. B., Forstner, W., Plumer, L., Steinhage, V., 1998. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Computer Vision and Image Understanding*, 72(2), pp. 185-203.
- Huertas, A., Nevatia, R., 1988. Detecting buildings in aerial images. *Computer Vision, Graphics and Image Processing*, 41(2), pp. 131-152.
- Lin, C. A., Nevatia, R., 1998. Building detection and description from a single intensity image. *Computer Vision and Image Understanding*, 72(2), pp. 101-121.
- Jaynes, C., Stolle, F., Collins, R., 1994. Task driven perceptual organization for extraction of rooftop polygons. *IEEE Workshop on Application of Computer Vision*, pp. 152-159.
- Noronha, S., Nevatia, R., 2001. Detection and modeling of buildings from multiple aerial images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(5), pp. 501-518.
- Mohan, R., Nevatia, R., 1989. Using perceptual organization to extract 3D structure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(11), pp. 1121-1139.
- Zabih, R., Woodfill, J., 1998. Non-Parametric Local Transforms for Computing Visual Correspondence. *Proc. Third European Conf. Computer Vision*, pp. 5-28.