Bidimensional Motion Charge Map for Stereovision Disparity Analysis

José M. López-Valles
Departamento de Telecomunicación
EUPC, Universidad de Castilla-La Mancha
13071 - Cuenca, Spain
JoseMaria.Lopez@uclm.es

Miguel A. Fernández
Antonio Fernández-Caballero
Departamento de Informática
EPSA, Universidad de Castilla-La Mancha
02071 - Albacete, Spain

Abstract

Up to date several strategies of how to retrieve disparity information from a sequence of images have been described. In this paper we introduce a method to retrieve disparity based on motion and stereovision. A motion representation in form of a bidimensional motion charge map, based in the so-called permanency memories mechanism is presented. For each pair of frames of a video stereovision sequence, the method displaces the left permanency stereomemory on the epipolar restriction basis over the right one, in order to analyze the disparities of the motion trails calculated.

1. Introduction

In general there are several strategies of how to retrieve depth information from a sequence of images, like depth from motion, depth from shading and depth from stereovision. In this paper we introduce a new method to retrieve disparity based on motion and stereovision. In a conventional stereoscopic approach, usually two cameras are mounted with a horizontal distance between them. As a consequence, objects displaced in depth from the fixation point are projected onto image regions which are shifted with respect to the image center. Due to the geometry of the optic system, it is sufficient to restrict disparity analysis to the projection of corresponding linear segments in the left and right camera. In some approaches, the disparity is computed by searching the maximum of the cross-correlation between image windows along the epipolar lines of the left and right image [11].

So far, many algorithms have been developed to analyze the depth in a scene. Brown et al. [2] describe a good approximation to all of them in their survey article. In many previous works, a series of restrictions are used to approach the correspondence problem. The most usual restriction is the disparity restriction, which considers that is not probable that there exist objects very close to the camera. The scene uses to be limited to a medium distance. This way, too high disparities are eliminated [14]. Koenderink and van Doorn [12] expressed the necessary theory in best initial works related to disparity restriction, and Wildes [16] implemented some of their ideas [17]. More recently, disparity in stereoscopy continues showing its great interest (e.g., [13], [1]).

According to the correspondence techniques used, we may classify methods into correlation-based, relaxation-based, gradient-based, and feature-based. The main correlation-based technique is the area correlation technique (e.g., [18]. The basic idea of relaxation techniques is that pixels to be set into correspondence perform ”controlled estimations”. In this kind of process, the correlation values of the neighbors of a pixel are of a great importance for the evaluation of the correspondence [10]. Methods based in the gradient or in the optical flow aim to determine local disparities between two images by formulating a differential equation that relates motion and luminance [3]. Techniques based in features limit to reliable features, such as contours or curves (e.g., [15]), at the analyzed regions.

Most methods have as a common denominator that they work with static images and not with motion information. In this paper, we have chosen as an alternative not to use direct information from the image, but rather the one derived from motion analysis. The system proposed uses as input the motion information of the objects present in the stereo scene, and uses this information to perform a depth analysis of the scene, through the use of a bidimensional motion charge map.

2. Bidimensional Motion Charge Map

The input to our system is a pair of stereo image sequences. These sequences have been acquired by means of two cameras arranged in a parallel configuration. The central idea behind our approach is to transpose the spatially-defined problem of disparity estimation into the temporal domain and compute the disparity simultaneously with the incoming data. This can be achieved realizing that in a well-calibrated fronto-parallel camera arrangement the epipolar lines are horizontal and thereby identical to the camera scan-lines. Thus, they will capture two similar, although not exactly equal, scenes. In case the images have been acquired in a convergent configuration, horizontal epipolar lines can be obtained by image-rectification techniques [4].

The motion analysis algorithm used in this work has already been tested in applications such as moving object shape recognition in noisy environments [6] [9], moving objects classification by motion features such as velocity or acceleration [5], and in applications related to selective visual attention [8]. Motion analysis performs separately on both stereovision sequences in two phases. The first analysis phase is based in grouping neighboring pixels that have similar grey levels in closed and connected regions in an image frame (of both stereo sequences). The method used is segmentation in grey level bands. This method consists in reducing the resolution of illumination levels of the image, obtaining this way a lower number of image regions, which potentially belong to a single object in motion. Let $B(x, y, t)$ be the grey level band associated to pixel (x, y) at time instant t, $GL(x, y, t)$ the grey level, n the number of grey level bands, and N the number of grey levels, then:

$$B(x, y, t) = \lfloor \frac{B(x, y, t-1) \cdot n}{N} + 0.5 \rfloor \tag{1}$$

A detailed analysis of the features and performances of this segmentation method is described in [7]. Obviously, segmentation in grey level bands performs in parallel on each couple of images of the stereo sequence.

Once the objects present in the scene are approximated in a broad way, the second phase has to detect possible motions of the segmented regions. Again, motion information of both video sequences that form the stereo pair is extracted. Motion detection is obtained from image pixels change in luminosity as the video sequence goes on through time. Motion in an image segmented in grey level bands is detected through the variation of the grey level band of the pixels. Notice that it is not that important that regions neither completely adjusts to the shape of the objects, nor that at a given moment two different objects appear overlapped in a same region. Consider that the proper relative motion of the objects will force those regions belonging to a same object to move in a uniform way, and those regions that hold different objects separate in the future.

From motion detection, we now introduce a representation that may help to establish further correspondences between different motion information. This representation finds its basis in the permanency memories mechanism. Precisely, this mechanism considers the jumps of pixels between bands, and it consists in a matrix of charge accumulators. The matrix, also called bidimensional motion charge map, is composed of as many units in horizontal and vertical direction as pixels there are in an image frame. This way, a position (x, y) of the image is associated to a permanency memory charge unit. Initially all accumulators are empty; that is to say, their charge is minimal. The charge in the permanency memory depends on the difference between the current and the previous images grey level band value. An accumulator detects differences $(diff(x, y, t)$ between the grey level bands of a pixel in the current and the previous frame:

$$diff(x, y, t) = \begin{cases} 0, & \text{if } B(x, y, t) = B(x, y, t-1) \\ 1, & \text{if } B(x, y, t) \neq B(x, y, t-1) \end{cases} \tag{2}$$

When a jump between grey level bands occurs at a pixel, the charge unit (accumulator) of the motion charge map at the pixel's position - $Ch(x, y, t)$ - is completely charged (charged to the maximum charge value max). This is the way to record that motion has just been detected at this pixel. This complete charge is produced when there is a jump to superior bands as well as to inferior bands. Thus, charge units of the permanency memory are able to inform on the presence of motion of the associated pixels. After the complete charge, each unit of the bidimensional motion charge map memory goes decrementing with time (in a frame by frame basis) down to reaching the minimum charge value min, while no motion is detected, or it is completely recharged, if motion is detected again.

This behavior is described by means of the following formula, where again $B(x, y, t)$ is the grey level band associated to pixel (x, y) at time instant t. dec is a fixed application-dependent quantity, which is decremented to the instantaneous charge of each charge unit each time that a frame is analyzed and no motion is detected. Thus, this quantity shows the discharge velocity of the permanency memory.

$$Ch(x, y, t) = \begin{cases} max, \\ \text{ if } diff(x, y, t) = 1 \\ \max[Ch(x, y, t-1) - dec, min], \\ \text{ if } diff(x, y, t) = 0 \end{cases} \tag{3}$$

Values of parameters dec, max and min have to be fixed according to the applications characteristics. Concretely, values max and min have to be chosen by taking into account that charge values will always be between them. dec defines the charge decrement interval between time instants $t-1$ and t. Thus, notice that the two-dimensional motion charge map stores motion information as a quantified value, which may be used for several classification purposes.

Thus, obviously, the evolution of charge in space depends on the velocity of the mobile in a direction. A slow mobile causes a short charge slope, as the object's advance from pixel to pixel may last various frames. During this

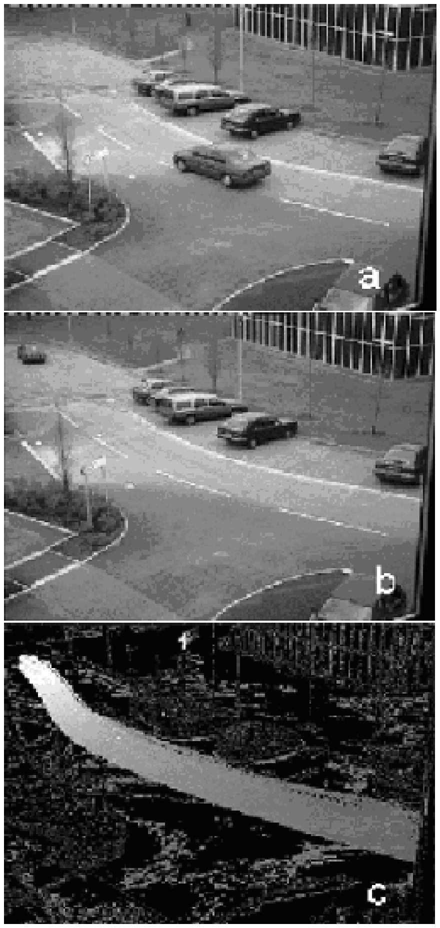

Figure 1. Motion charge map: (a) one image of a sequence, (b) same perspective after some seconds, (c) motion trails as represented on the bidimensional motion charge map

time elapsed all affected units are discharging. In this case, between the charge and discharge of a unit, the mobile covers a short distance. On the other hand, a quick mobile causes various memory units to charge simultaneously, such that there many more units will be affected by this motion. In this second case, between the total charge and discharge of a unit of the memory the mobile covers many pixels. Figure 1 shows all these issues. Figure 1a and Figure 1b show two images of a monocular sequence. The advance of a car may be appreciated, as well as a more slight movement of a pedestrian. In Figure 1c you may observe the effect of these moving objects on the motion charge map.

The difference between a quick object as the car, which is leaving a very long motion trail (from dark grey to white), and a pedestrian whose velocity is clearly slower and whose motion trail is nearly unappreciable with respect to the cars one, is presented. Thus, motion charge maps enable representing the motion history of the frames that form the image sequence, that is to say, there is segmentation from the motion of the objects present in the scene.

However, the dependency of the permanency memories from the segmentation in grey level bands imposes a restriction. The diminishment of the resolution in illumination levels produced by the segmentation in grey level bands does not exactly imply segmentation into objects. Some of the objects of the images are segmented into various regions, and physically distinct objects may be overlapped into a same region. Nevertheless, this issue is not that important when taking into account that our aim is to characterize motion of the objects and not their shape.

3. Stereovision Disparity Analysis

Motion-based segmentation into a bidimensional motion charge map, as explained in the previous section, facilitates the correspondence analysis. Indeed, motion trails obtained through the permanency memories charge units are used to analyze the disparity between the objects in the stereo pair in a more easy and precise way. The set of all disparities between two images of a stereo pair is called the disparity map.

The retrieval of disparity information is usually a very early step in image analysis. It requires stereotyped processing where each single pixel enters the computation. In stereovision, methods based on local primitives as pixels and contours may be very efficient, but they are too much sensitive to locally ambiguous regions, such as occlusions or uniform texture regions. Methods based on areas are less sensitive to these problems, as they offer an additional support to obtain correspondences of difficult regions in a more easy and robust way, or they discard false disparities. Although methods based on areas use to be computationally very expensive, we introduce a simple area-based method with a low computational cost.

In order to explain our disparity analysis method, it is sufficient to analyze the process at the level of epipolar lines. The key idea is that a moving object causes two identical trails to appear in epipolar lines of the permanency stereo-memories - see pair of bidimensional motion charge maps. The only difference relies in their relative positions, affected by the disparity of the object at each moment.

In Figure 2, the charge values in two corresponding superimposed epipolar lines of the maps are represented. In a parallel configuration as the one we have chosen, there will be no disparity in right and left image for objects that are in a great depth - imagine in the infinite. Nevertheless, when an object approaches to the central point of the base line, that is to say, between the two cameras, the object goes appearing more to the right on the left image and more to the left on the right image. This is precisely the disparity con-

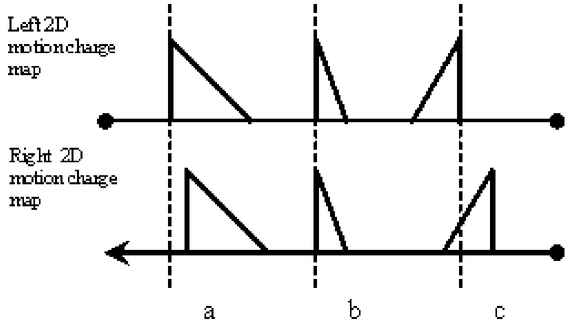

Figure 2. Disparity by motion charge maps

cept; the more close objects have a greater disparity than the more distant ones.

Looking at Figure 2 it is possible to analyze the motion of each one of the three objects present in the permanency memories from their motion trails. This initial analysis is independent of the epipolar constraint studied. You may observe that object ”a”, which has a long trail and has his maximum charge towards the left, is advancing to the left at a high speed. Object ”b”, with a shorter trail, is also advancing towards the same direction but at a slower velocity. Finally, object ”c”, whose trail is inverted in horizontal, is moving to the right at a medium velocity, as shown by its trail.

Also from Figure 2, but now comparing between the motion trails in both epipolar lines, disparity is analyzed. Motion trail of object ”b” presents a null disparity. Therefore, we can conclude that this trail corresponds to an object that is far away from the cameras. Remember that due to our parallel cameras configuration, pixels with a null disparity are located in the infinite. Object ”a” has a little greater disparity. Finally, object ”c” offers the greatest disparity.

This simple example draws three main conclusions. Firstly, in order to consider two motion trails to be correspondent, it must only be checked that both are equal enough in length and in discharge direction in epipolar lines of the pair of bidimensional motion charge maps. Secondly, we may affirm that, in order to analyze disparities, one possibility is to displace one epipolar line over the other one, until we get the exact point where both lines are completely superimposed. In other words, an epipolar line has to be displaced over the other until motion trails coincide. Of course, the right epipolar line can be displaced over the left or the left epipolar line over the right. When the motion trails coincide, the displacement value applied to the epipolar line is the disparity value. In third place, if we consider the representation of a mobile with a high velocity, various charge units of the permanence memories may charge simultaneously. This way, an object may correspond to various disparities. This is the reason why one single memory unit is not able to establish the disparity of an object. It is

Figure 3. Frame 211 of the "OutdoorZoom" stereo sequence. (Top Left) Original Left Image, (Top Right) Original Right Image, (Bottom Left) Rectified Left Image, (Bottom Right) Rectified Right Image

necessary to analyze the correspondence from the values of various units. The decision of all units has to validate the overall disparity value. The more efficient way to manage this is that each pixel chooses its disparity in such a way that the maximum of its neighboring units confirm the disparity.

All these considerations tell us that the disparity analysis at epipolar line level consists in superimposing both epipolar lines with different relative displacements and in analyzing the correspondences produced in the neighborhood of each unit. The one displacement, which produces that a maximum number of surrounding elements confirm its correspondence, demonstrates to be the more trustful disparity value.

4. Data and Results

In order to test our algorithms, a couple of real stereo sequences are shown. Firstly, we show the results of applying our algorithms to a scenario called ”OutdoorZoom” (see Figure 3), downloaded from labvisione.deis.unibo.it/ smattoccia/stereo.htm.

The whole sequence is 30 seconds long and has been acquired at a rate of 10 images per second. The values of the main parameters used in our test series were:

```
dec = 128; n = 8; min = 0; max = 255;
```

Figure 4 shows the result for some of the more representative results of applying our algorithms to the ”OutdoorZoom” scenario. In column (a) the segmentation in grey level bands may be appreciated, in column (b) motion information as represented in the right motion charge map is offered, and in column (c) the final output, that is to say, the scene depth as detected by the cameras, is presented.

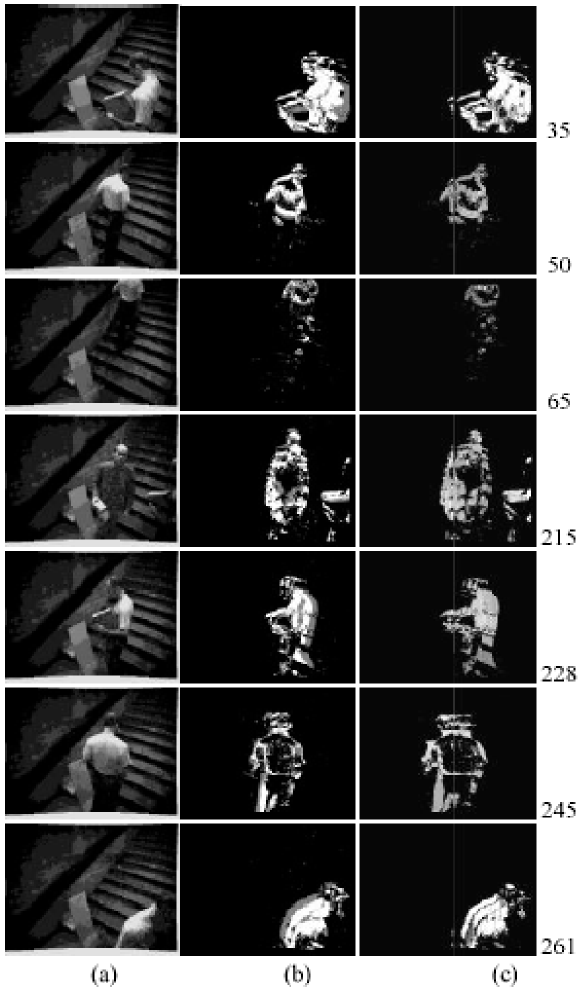

Figure 4. Results for "OutdoorZoom" scenario

You may observe on Figure 4 that light colors in column (c) means that persons are closer to the cameras. Black means there is no motion detected. The main information is available in columns (b) and (c). We may observe some details, as, for example, the following ones:

- In frame 35, a person is entering the scene on the right side, very close to the cameras. This is why, in column (c), the final output, very light grey levels appear.
- This person progressively is moving away from the cameras, in such a way that on frame 50 it is represented by intermediate grey levels.
- In frame 65, the person is now far away from the cameras. Its shape appears in dark grey values.
- Let us now focus on frame 215. A person is walking down the steps and at the same time an object is appearing on the right side of the image. It may be appreciated at the output of the system that the object is a bit lighter than the person. Thus, the object has to be closer to the cameras than the walking person.
- From frame 215 to frame 228, the pedestrian is walking horizontally (to the left). Thus, we appreciate no difference in the grey levels present in these frames.
- In frame 245, the person turns around, but there is still no difference appreciated in its depth in the scene.
- Lastly, in frame 261, we may observe the person leaving the scene on the right side, and at the output very light grey levels. This obviously means that the man is very close to the cameras.

Figure 5. Frame 53 of the "IndoorZoom" stereo sequence. (Top Left) Original Left Image, (Top Right) Original Right Image, (Bottom Left) Rectified Left Image, (Bottom Right) Rectified Right Image

In second place, we also show some results of the scenario called ”IndoorZoom” (see Figure 5), downloaded from labvisione.deis.unibo.it/ smattoccia/stereo.htm, as well. In this scenario, two cameras are situated over an entrance door.

The whole sequence is 29.9 seconds long and has been acquired at a rate of 10 images per second. Figure 6 shows some of the more representative results of applying our algorithms to the ”IndoorZoom” scenario.

The real interest of this series of images is related to occlusions. The motion trails, and moreover the depths, of people in the scene are different, enabling this way to distinguish among different persons.

5. Conclusions

In this paper we have introduced a new method to retrieve disparity information based on motion and stereovision. A motion detection representation helps establishing

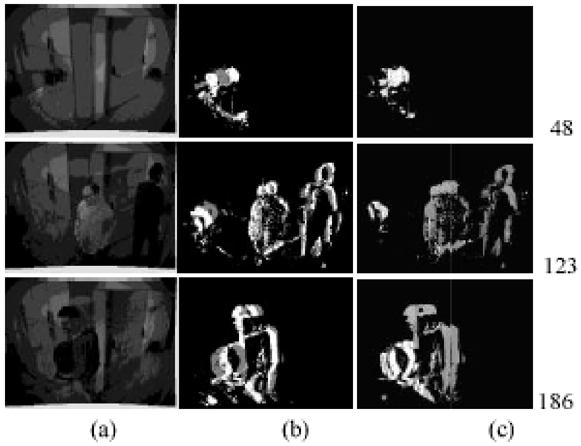

Figure 6. Results for "IndoorZoom" scenario

further correspondences between different motion information. This representation bases in the permanency memories mechanism, where jumps of pixels between grey level bands are computed in a matrix of charge accumulators. Thus, for the purpose to analyze scene depth from stereo images, we have chosen the alternative not to use direct information from the image, but rather the one derived from motion analysis. This alternative provides an important advantage.

Trough motion information stored in bidimensional motion charge maps it is easier to use correspondences than by grey level information of the frames. The results are also more accurate and robust. This is due to the instantaneous motion features, such as position, velocity, acceleration and direction of the diverse moving objects. Motion information of an object is different from any other moving object's one. Nonetheless, when observing motion features of a concrete object in both stereo sequences at the same time instant, we appreciate that these features are extremely similar. This is the reason why it is easy and robust to establish correspondences between the motion information of an object at the right image respect to the object at the left image. There exist less ambiguity possibilities.

6. Acknowledgement

This work is supported in part by the Spanish CICYT TIN2004-07661-C02-02 grant.

References

[1] S. G. andf J.L. Marroquin. Robust Approach for Disparity Estimation in Stereo Vision. *Image and Vision Computing*, 22(3):183 – 195, 2004.

[2] M. Brown, D. Burschka, and G. Hager. Advances in Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993 – 1008, 2003.

[3] I. Choi, J.-G. Yoon, Y.-B. Lee, and S. Chien. Stereo System for Tracking Moving Objects using Log-polar Transformation and Zero Disparity Filtering. In *CAIP*, pages 182 – 189, 2003.

[4] O. Faugeras. *Three-dimensional Computer Vision*. The MIT Press, 1993.

[5] M. Fernández, A. Fernández-Caballero, M. López, and J. Mira. Length-speed Ratio as a Characteristic for Moving Elements Real-time Classification. *Real-Tiem Imaging*, 9:49 – 59, 2003.

[6] A. Fernández-Caballero, M. Fernández, J. Mira, and A. Delgado. Spatio-temporal Shape Building from Image Sequences using Lateral Interaction in Accumulative Computation. *Pattern Recognition*, 36(5):1131 – 1142, 2003.

[7] A. Fernández-Caballero, M. López, M. Fernández, J. Mira, A. Delgado, and J. López-Valles. Accumulative Computation Method for Motion Features Extraction in Dynamic Selective Visual Attention. In *Proc. 2nd International Workshop on Attention and Performance in Computational Vision*, 2004.

[8] A. Fernández-Caballero, J. Mira, A. Delgado, and M. Fernández. Lateral Interaction in Accumulative Computation: A Model for Motion Detection. *Neurocomputing*, 50:341 – 364, 2003.

[9] A. Fernández-Caballero, J. Mira, M. Fernández, and A. Delgado. On Motion Detection through a Multi-layer Neural Network Architecture. *Neural Networks*, 16(2):205 – 222, 2003.

[10] W. Grimson. Computational Experiments with a Feature based Stereo Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:17 – 34, 1985.

[11] R. Haralick and L. Shapiro. *Computer and Robot Vision*. Addison-Wesley, 1992.

[12] J. Koenderink and A. van Doorn. Geometry of Binocular Vision and a Model for Stereovision. *Biological Cybernetics*, 21:29 – 35, 1976.

[13] K. Muhlmann, D. Maier, J. Hesser, and R. Manner. Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation. *International Journal of Computer Vision*, 47(1–3):79 – 88, 2002.

[14] Y. Sumi, Y. Kawai, T. Yoshimi, and F. Tomita. 3d Object Recognition in Cluttered Environments by Segment-based Stereo Vision. *International Journal of Computer Vision*, 46(1):5 – 23, 2002.

[15] V. Venkateswar and R. Chellappa. Hierarchical Stereo and Motion Correspondence using Feature Groupings. *International Journal of Computer Vision*, 15:245 – 269, 1995.

[16] R. Wildes. Direct Recovery of Three-dimensional Scene Geometry from Binocular Stereo Disparity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):761 – 774, 1991.

[17] R. Wilson and H. Knutsson. A Multiresolution Stereovision Algorithm based on Gabor Representation. In *Proc. IEE International Conference on Image Processing and Applications*, pages 19 – 22, 1989.

[18] R. Zabih and J. Woodfill. Non Parametric Local Transforms for Computing Visual Correspondence. In *Proc. 3rd European Conference on Computer Vision*, pages 150 – 158, 1994.