Multispectral Stereo Image Correspondence

Marcelo D. Pistarelli[1], Angel D. Sappa[2], and Ricardo Toledo[2]

[1] Departamento de Sistemas e Informática
Facultad de Cs. Exactas, Ingeniería y Agrimensura
Universidad Nacional de Rosario, Rosario, Argentina
[2] Computer Vision Center, Edifici O, Campus UAB,
08193 Bellaterra, Barcelona, Spain
mpistare@fceia.unr.edu.ar, {asappa,ricard}@cvc.uab.es

Abstract. This paper presents a novel multispectral stereo image correspondence approach. It is evaluated using a stereo rig constructed with a visible spectrum camera and a long wave infrared spectrum camera. The novelty of the proposed approach lies on the usage of Hough space as a correspondence search domain. In this way it avoids searching for correspondence in the original multispectral image domains, where information is low correlated, and a common domain is used. The proposed approach is intended to be used in outdoor urban scenarios, where images contain large amount of edges. These edges are used as distinctive characteristics for the matching in the Hough space. Experimental results are provided showing the validity of the proposed approach.

1 Introduction

Multispectral image processing has been largely studied by the remote sensing community where images from different spectral bands are generally used to extract information related with the characteristics of the materials defining the different regions of the image (e.g., vegetation, water, etc.). The stereo matching problem, although out of the main scope of remote sensing community, has been also tackled in the context of remote sensing [1].

Recent advances in multispectral imaging sensors have opened new opportunities to the research community to tackle classical computer vision problems, which are not related with the remote sensing field. Actually, nowadays multispectral applications can be found in: video surveillance (e.g., [2], [3]), driver assistance [4], energy saving [5], fire detection, etc. In most of these applications finding correspondences, or matching, between features observed in the different images is the first problem to be tackled. Feature matching is a common task in both the image registration and the stereovision problems. In both cases feature matching has been tackled for the case where the images belong to the same spectral band. In general, the approaches in the literature are based on key points detection and their corresponding description. The descriptions of such key points are later on used for finding the correspondences (e.g., [6], [7]). This kind of scheme has been recently used for finding correspondences in the

R. Wilson et al. (Eds.): CAIP 2013, Part II, LNCS 8048, pp. 217–224, 2013.

multispectral domain (e.g., [8] and [9]), but due to the nonlinear relationship between pixel intensities the matching ratio remains low and the usage of scene prior is needed (e.g., [10], [11]).

Focussing on the stereo matching problem, the current work proposes a novel approach that formulates the multispectral matching problem by representing the given image pair in a common space (Hough space). Actually, in this common space not all the image pixels are represented but just those ones corresponding to the edges extracted with the Canny algorithm [12]. The proposed approach consists of three stages. First, the edges from the multispectral images are extracted using an adaptive threshold strategy that allows to generate similar representations from the image pair. Then, the edges are represented in the common space using the classical Hough transform [13] that identifies straight lines. Finally, the most representative lines are extracted and their parameters used as descriptors to find the matchings. The proposed approach has been evaluated using a multispectral stereo rig constructed with a Long-Wave Infrared (LWIR) camera and a Visible Spectrum (VS) camera.

The manuscript is organized as follow. State of the art on multispectral stereo image correspondence problem is introduced in Section 2. The proposed approach is then presented in Section 3. Experimental results are provided in Section 4. Finally, conclusions are given in Section 5.

2 Related Work

Stereo vision problem involves different tasks, from the synchronized acquisition of a pair of images, till dense disparity maps and 3D representations. In between, tasks such as camera calibration, image rectification, correspondence search, and triangulation should be tackled. There is a large literature on all these topics (e.g., [14],[15]) when the cameras in the stereo rig work at the same spectral band. However, both the coexistence of cameras working at different spectral band and the needs to extract 3D information from such a multispectral devices require some reformulation in the pipeline of tasks mentioned above. In the current work a particular case where the cameras in the stereo rig work at the visible and infrared spectrum is considered. From the list of tasks mentioned above, we just consider the correspondence search problem assuming the multispectral stereo pair is already calibrated and rectified. Hence, epipolar lines correspond to the same rows in the different pair of images.

Finding correspondences in this multispectral scheme is a difficult task that has been solved in the literature using different strategies. For instance, the lack of overlap between VS (0.4-0.7 μm) and LWIR (9-15 μm) spectrum has been tackled in [16] through a cost function based on mutual information that is enriched with gradient information in a scale space representation. The usage of this cost function allows to obtain sparse stereo representations that are later on filled in with scene prior information generating a dense disparity map [11].

On the contrary to the previous approaches, the authors in [3] constraint the matching to those regions that contain human body silhouettes. Since their contribution is aimed at person tracking, some assumptions are applied, for example

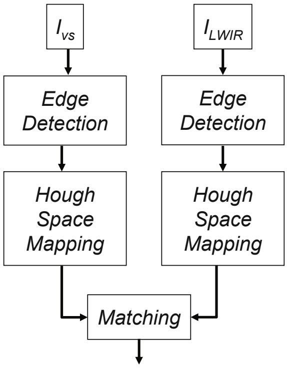

Fig. 1. Pipeline of the proposed approach (I_{VS}: visible spectum image; I_{LWIR}: long-wave infrared image)

a foreground segmentation for disclosing possible human shapes, which are corresponded by maximizing mutual information. Although, these assumptions are valid, they restrict the scope of applications to those scenarios containing pedestrians. Furthermore, it should be noted that in [3] only 3D information on those pixels defining the surface of the pedestrians body is extracted.

Inspired by the classical SIFT algorithm [6] a descriptor based on the edge histogram is proposed in [8] to find correspondences between VS and LWIR images. The proposed edge oriented histogram is then used to describe points of interest detected in a scale space representation. A scale restriction criteria is used in [9] to reduce the amount of mismatching of SIFT when it is adopted to tackle the VS and NIR correspondence case. In this particular multispectral case the spectral bands of the pair of images are nearer than in the VS-LWIR case, hence the number of right matching is considerably higher in this case.

3 Proposed Method

We propose a novel approach for finding correspondences between multispectral image pairs. Actually, only edges are considered, by representing them in a common space (Hough space). The proposed approach consists of three stages (see the pipeline in Fig. 1): *i*) edge detection; *ii*) Hough space mapping; *iii*) matching; these stages are detailed below. A case study (see Fig. 2) is used to illustrate the three stages.

3.1 Edge Detection

Since the proposed approach is intended for finding correspondences in multispectral stereo images from urban scenarios, man-made structures are the predominant features in the given images. The shape of these structures can be in part captured through an edge detector algorithm. In the current work the

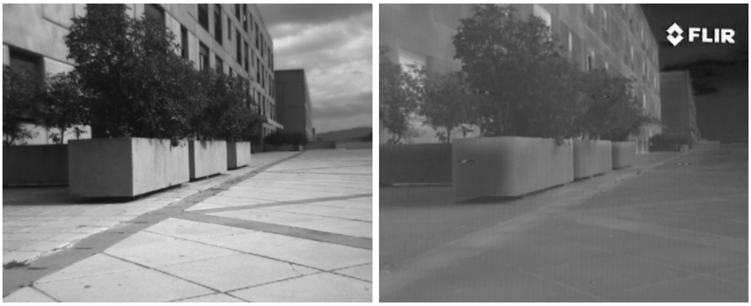

Fig. 2. Multispectral stereo pair: (*left*) I_{VS} image; (*right*) I_{LWIR} image

Canny algorithm [12] is used and a scheme that automatically adapt the algorithm' thresholds is considered. This scheme allows to generate images with a similar amount of edges. Figure 3 shows the binary images $B(x, y)$ of edges extracted from a multispectral pair using the same algorithm setup in both cases. It can bee appreciated that there is a difference on the amount of extracted edges, which at the end will affect their representation in the Hough space.

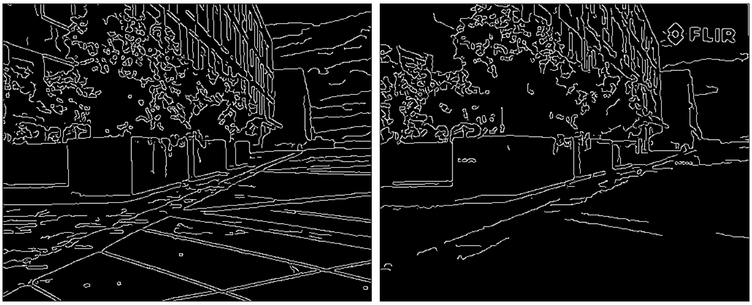

Fig. 3. Edges extracted using the Canny algorithm with the same parameters in both cases: (*left*) VS image, (*right*) LWIR image

In order to diminish the difference in the amount of edges mentioned above an iterative approach is performed over one of the images in order to get a similar amount of edges in both cases. During the iterations the Canny's thresholds are updated. Finally, in order to reduce the amount of edges that do not contribute to the proposed scheme, which consist in detecting the most representative geometric features in the Hough space, edges defined by less than T_{ha} pixels are removed (in the current implementation $T_{ha} = 20$). Results from this adaptive threshold and short edge filtering are presented in Fig. 4. Although the amount of edges in this pair of images are different they look more similar than the one presented in Fig. 3.

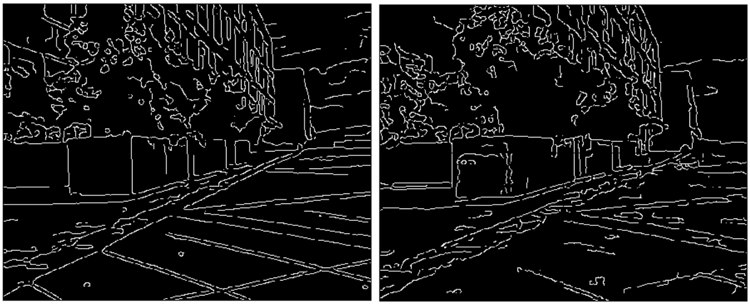

Fig. 4. (*left*) Edges for the VS image with a user defined threshold. (*right*) Edges for the LWIR image extracted from a the Canny algorithm with an adaptive threshold

3.2 Hough Space Mapping

The Hough transform [13] for a binary image $B(x, y)$, defined as:

$$H(\theta, \rho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(x, y)\delta(\rho - x\cos\theta - y\sin\theta)dxdy \tag{1}$$

transform every pixel from B into a sine wave $\rho = x\cos\theta - y\sin\theta$. Figure 5 shows an illustration of the mapping of two image points to the Hough space. The intersection point in the Hough space corresponds to the parameters of the straight line passing through (x, y), (u, v) in the image space.

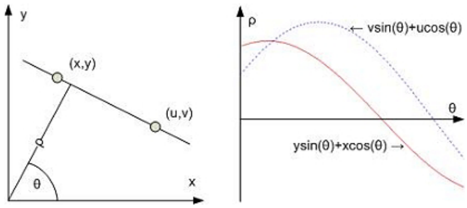

Fig. 5. Mapping of two image points to the Hough space

By applying the Hough transform to all the pixels defining the edges extracted in the previous stage we obtain a set of sine wave overlapped in the Hough space. Hence, by accumulating these mappings into a $N \times M$ matrix, the most representative ones can be easily identified just by looking at the cells with larger amounts of votes. Figure 6 depicts the Hough representations corresponding to the edges presented in Fig. 4.

3.3 Matching

The Hough representations obtained above can be understood as the feature description space that will be used for finding the matchings between the multispectral images. Essentially, this stage consists of matching points from the

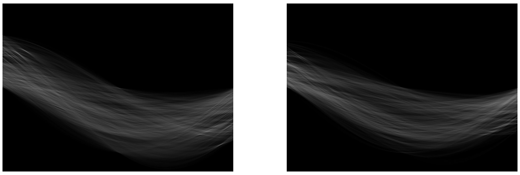

Fig. 6. Hough representations of edges from Fig. 4 (*left*) and edges from Fig. 4 (*right*)

Hough representations that have large amount of votes and the distance between them is smaller that a user defined threshold. Additionally, a non-maximum suppression algorithm is used to avoid wrong associations. Once a matching between two points in the Hough spaces has been found (a matching is accepted when a similar orientation in a neighborhood is found, assuming a disparity value of $\pm\Delta$) the corresponding edge points in the image spaces are extracted by finding the intersections between epipolar lines and straight lines from Hough. Note that since the stereo pair is rectified, epipolar lines correspond with horizontal lines in the image pair. Figure 7 presents the matchings obtained in the multispectral pair used as a case study through the previous sections. Other more elaborated strategies for finding matchings could be used, for instance dynamic programming [17] or graph cut [18] based approaches, which would improve the results but increasing considerably the processing time.

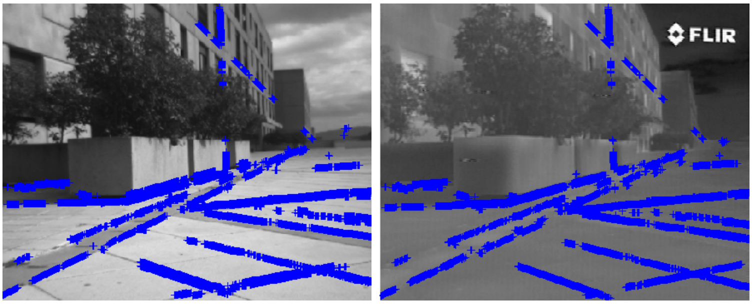

Fig. 7. Correspondences (5163 points) obtained with the proposed approach

4 Experimental Results

The proposed approach has been evaluated with a set of rectified multispectral stereo images [8] showing the validity of the proposed approach. In all the cases the number of matchings obtained with the proposed approach are considerable larger than those presented in [8]. Note that points matched with the proposed approach lie over image edges, which is one of the difference with respect to previous approaches where key points are used for finding correspondences, instead of higher level features. Figure 8 shows two pairs of multispectral images with the points matched with the proposed approach. On average, 100 times more points are correctly matched with the proposed approach than [8] when the same data set is considered (the data set contains 100 pairs of multispectral images and is available through the following link: www.cvc.uab.es/adas/projects/simeve).

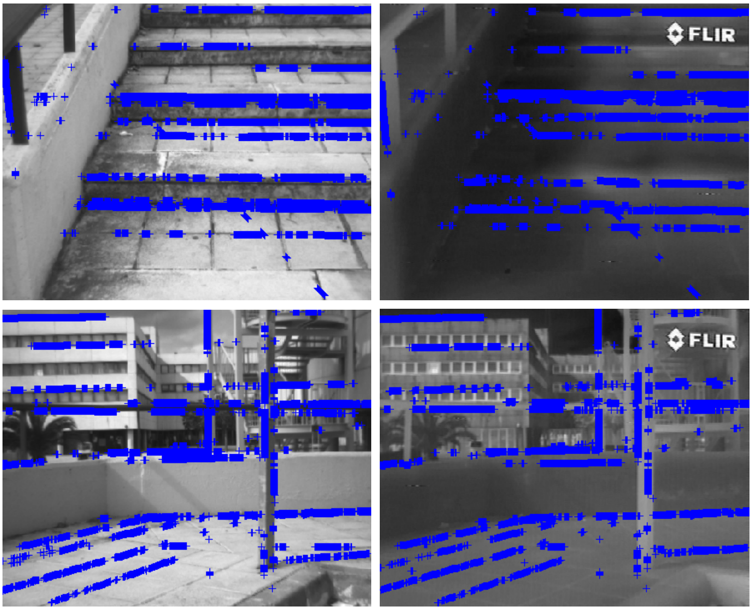

Fig. 8. Correspondences obtained with the proposed approach: (*top*) 5447 points; (*bottom*) 4512 points

5 Conclusions

This paper presents a novel approach that overcome the problem of finding feature point correspondences when images from different spectral bands are considered. The proposed solution consists in representing the given images in a common space. It is mainly intended to tackle outdoor urban scenarios, where straight lines are usually the most predominant shape characteristics. Future work will be focussed on the usage of a scale space scheme to increase the number of correct matchings as well as on the usage of the obtained correspondences towards a dense disparity map.

Acknowledgments. This work has been partially supported by the Spanish Government under Research Project TIN2011-25606.

References

1. Felicísimo, A., Cuartero, A.: Methodological proposal for multispectral stereo matching. IEEE Trans. on Geoscience and Remote Sensing 44, 2534–2538 (2006)
2. Leykin, A., Hammoud, R.: Pedestrian tracking by fusion of thermal-visible surveillance videos. Machine Vision and Applications 21, 587–595 (2010)

3. Krotosky, S., Trivedi, M.: On color-, infrared-, and multimodal-stereo approaches to pedestrian detection. IEEE Trans. on Intelligent Transportation Systems 8, 619–629 (2007)
4. Jung, S., Eledath, J., Johansson, S., Mathevon, V.: Egomotion estimation in monocular infra-red image sequence for night vision applications. In: IEEE Workshop on Applications of Computer Vision, Austin, TX, USA, February 20-21 (2007)
5. Prakash, S., Lee, P.Y., Caelli, T.: 3d mapping of surface temperature using thermal stereo. In: Int'l Conf. on Control, Automation, Robotics and Vision, pp. 1–4 (2006)
6. Lowe, D.: Object recognition from local scale-invariant features. In: IEEE Conference on Computer Vision, Kerkyra, Greece, pp. 1150–1157 (1999)
7. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
8. Aguilera, C., Barrera, F., Lumbreras, F., Sappa, A., Toledo, R.: Multispectral image feature points. Sensors 12, 12661–12672 (2012)
9. Yi, Z., Zhiguo, C., Yang, X.: Multi-spectral remote image registration based on sift. Electronics Letters 44, 107–108 (2008)
10. Barrera, F., Lumbreras, F., Sappa, A.: Multimodal stereo vision system: 3d data extraction and algorithm evaluation. IEEE Journal of Selected Topics in Signal Processing 6, 437–446 (2012)
11. Barrera, F., Lumbreras, F., Sappa, A.: Multispectral piecewise planar stereo using manhattan-world assumption. Pattern Recognition Letters 34, 52–61 (2013)
12. Canny, J.: A computational approach to edge detection. IEEE Trans. on Pattern Analysis and Machine Intelligence 8, 679–698 (1986)
13. Hough, P.: Method and means for recognizzing complex patterns. U.S. Patent 3,069,654 (1962)
14. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2004)
15. Faugeras, O., Luong, Q.: The Geometry of Multiple Images: the Laws that Govern the Formation of Multiple Images of a scene and Some of their Applications. The MIT Press (2001)
16. Barrera, F., Lumbreras, F., Sappa, A.: Multimodal template matching based on gradient and mutual information using scale-space. In: IEEE Int'l Conf. on Image Processing, Hong Kong, pp. 2749–2752 (2010)
17. Li, Z.N.: Stereo correspondence based on line matching in Hough space using dynamic programming. IEEE Trans. on Systems, Man, and Cybernetics 24, 144–152 (1994)
18. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. on Pattern Analysis and Machine Intelligence 23, 1222–1239 (2001)