

Infrared-Visual Image Registration Based on Corners and Hausdorff Distance^{*}

Tomislav Hrkać, Zoran Kalafatić, and Josip Krapac

University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
tomislav.hrkac@fer.hr

Abstract. The paper presents an approach to multimodal image registration. The method is developed for aligning infrared (IR) and visual (RGB) images of facades. It is based on mapping clouds of points extracted by a corner detector applied to both images. The experiments show that corners are suitable features for our application. In the alignment process a number of transformation hypotheses is generated and evaluated. The evaluation is performed by measuring similarity between the RGB corners and the transformed corners from IR image. Directed partial Hausdorff distance is used as a robust similarity measure. The implemented system has been tested on various IR-RGB pairs of images of buildings. The results show that the method can be used for image registration, but also expose some typical problems.

1 Introduction

Registration is a fundamental problem in computer vision. The main task is to align two images taken at different times, from different viewpoints, or by different sensors. There are several comprehensive reviews of the subject [1,2].

When developing an image registration algorithm, several key issues have to be addressed:

- appropriate image transformation class for aligning one image to another;
- features to be used as landmarks guiding the registration procedure;
- strategy for hypothesis generation;
- similarity measure used for hypothesis evaluation.

In the case of multisensor registration, the images to be aligned are not necessarily similar, so that appearance-based or correlation-based registration methods cannot be used. It is necessary to use features that are stable with respect to sensor, i.e., the same physical artifact produces features in both images. Usual features used for multisensor registration are edges, line segments [3], or virtual line intersections [4,5]. Corners can also be used, but it seems that they are

^{*} This work has been supported by the Croatian Ministry of Science, Education and Sports, as a part of the TEST (technological R&D) programme, administrative number #4046 (2004).

not very popular due to their sensitivity to scale, skew, rotation, illumination changes, etc. [4].

This work is aimed to aligning infrared (IR) and visual (RGB) images of facades in an application for thermal isolation inspection. Experiments with various features extracted from IR and RGB images of buildings have shown that corners have the desired property of being stable in both images, due to the typical scene structure containing windows, doors, balconies, and similar rectangular areas (Fig. 1). As a typical scene contains mostly planar surfaces and both images (IR and RGB) are taken from approximately the same position and frontally to the facade, a simple similarity transform can be used for aligning the images. The strategy for hypothesis generation is based on establishing correspondences between pairs of corners from IR image and corner pairs in RGB image. Each hypothesis is evaluated by computing the corresponding similarity measure based on partial Hausdorff distance [6] between the transformed corners from IR image and the corners in RGB image. This similarity measure is not based on the correlation between the detected corners' neighbourhoods [7], but on the constellation of the corner positions.



Fig. 1. A typical RGB-IR image pair

The rest of the paper is organized as follows. The specific assumptions of the proposed approach to the multi-sensor image registration are described in Section 2. In Section 3 the registration procedure is presented, together with some implementation details. Section 4 presents and discusses the results of applying the implemented method. Some concluding remarks and a discussion of future research are given in Section 5.

2 Registration Assumptions

The proposed registration algorithm is tailored to the registration of IR and RGB images of facades. The nature of the application enables us to impose several constraints and simplify the algorithm design.

The first constraint is due to the fact that the images are usually taken frontally to the facade and the photographer tries to take the pictures of the

same part of the building. Since both images are taken from viewpoints that are not far apart, the general transform for registration of planar scenes (planar homography) can be approximated by much simpler similarity transform. The transformation is defined by four parameters: scaling (s), rotation (α), and translation (t_x, t_y):

$$\begin{bmatrix} x_{RGB} \\ y_{RGB} \end{bmatrix} = s \cdot \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \cdot \begin{bmatrix} x_{IR} \\ y_{IR} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \tag{1}$$

Assuming that the geometric transformation between the two images is global, two pairs of corresponding points are enough to compute the transformation parameters. As the structure of the images is such that it promotes corners in both the IR and RGB images, hypotheses are generated based on points obtained by a corner detector. The hypothesis generation procedure is discussed in Sec. 3.

Two corner detectors were tried out: SUSAN [8] and Harris [9]. In the conducted experiments, Harris corner detector has given significantly better results. That is consistent with [10]. An example of corners detected by Harris corner detector is shown in Fig. 2.



Fig. 2. Corners detected in IR (left) and RGB (right) images by using Harris corner detector

The Harris corner detector [9] determines the matrix \mathbf{C} describing the intensity structure of the local neighborhood W for each image pixel (Eq. 2):

$$\mathbf{C} = \begin{bmatrix} \sum_W (I_x(x, y))^2 & \sum_W (I_x(x, y) \cdot I_y(x, y)) \\ \sum_W (I_x(x, y) \cdot I_y(x, y)) & \sum_W (I_y(x, y))^2 \end{bmatrix}, \tag{2}$$

where $I_x(x, y) = \partial I(x, y) / \partial x$ and $I_y(x, y) = \partial I(x, y) / \partial y$. The detector output (corner “strength”) is computed as $s = \det(\mathbf{C}) - \kappa \cdot \text{trace}^2(\mathbf{C})$. Local maxima of these corner “strengths” indicate potential corner positions. In the experiments the parameter κ was set to 0.04. In order to prune the list of corners, local maxima having $s < 0.01 \cdot s_{max}$ were filtered out, where s_{max} denotes the strength of the global maximum. Since the IR images had significantly lower contrast than the RGB images, the contrast of IR images was enhanced by histogram equalization.

3 Hypothesis Generation and Evaluation

In order to compute the four transformation parameters (s, α, t_x, t_y) , a pair of points in the IR image and the corresponding pair of points in the RGB image must be known, providing a set of four equations. If the two corner points in the IR image are denoted by $A_1(x_{A1}, y_{A1})$ and $A_2(x_{A2}, y_{A2})$, and the two corresponding corner points in the RGB image by $B_1(x_{B1}, y_{B1})$ and $B_2(x_{B2}, y_{B2})$ (where point B_1 in the RGB image corresponds to the point A_1 in the IR image, and point B_2 corresponds to the point A_2), we get the following set of equations:

$$\begin{aligned} x_{B1} &= s \cdot (\cos \alpha \cdot x_{A1} - \sin \alpha \cdot y_{A1}) + t_x, \\ y_{B1} &= s \cdot (\sin \alpha \cdot x_{A1} + \cos \alpha \cdot y_{A1}) + t_y, \\ x_{B2} &= s \cdot (\cos \alpha \cdot x_{A2} - \sin \alpha \cdot y_{A2}) + t_x, \\ y_{B2} &= s \cdot (\sin \alpha \cdot x_{A2} + \cos \alpha \cdot y_{A2}) + t_y. \end{aligned} \quad (3)$$

By solving the above set of equations, the parameters of the transformation can be found as:

$$\alpha = \operatorname{arctg} \left(\frac{\Delta x_A \cdot \Delta y_B - \Delta x_B \cdot \Delta y_A}{\Delta y_A \cdot \Delta y_B + \Delta x_B \cdot \Delta x_A} \right),$$

$$s = \frac{\Delta x_B}{\cos \alpha \cdot \Delta x_A - \sin \alpha \cdot \Delta y_A}, \quad (4)$$

$$t_x = x_{B1} - s \cdot (\cos \alpha \cdot x_{A1} - \sin \alpha \cdot y_{A1}),$$

$$t_y = y_{B1} - s \cdot (\sin \alpha \cdot x_{A1} + \cos \alpha \cdot y_{A1}),$$

where $\Delta x_A = x_{A2} - x_{A1}$, $\Delta y_A = y_{A2} - y_{A1}$, $\Delta x_B = x_{B2} - x_{B1}$, and $\Delta y_B = y_{B2} - y_{B1}$.

The basic idea is to use all combinations of corner pairs from IR and RGB image as *match hypotheses*. For each match hypothesis, parameters of similarity transform are calculated according to equations (4), and each hypothesis is evaluated according to the criterion described later in this section. The transformation obtaining the best similarity measure is then selected as the final solution.

However, the described simple scenario has an obvious drawback: if *all* combinations of corner pairs were used, a very large number of hypotheses would be generated and evaluated. Many of those hypotheses are obviously false, which results in an unacceptably large computational load. Therefore, to achieve better computational times, the number of generated hypotheses has to be significantly reduced. Two strategies to reduce the number of hypotheses were employed.

First, because the images were taken from approximately the same viewpoint, the corresponding corners in the IR and RGB image should not be too far away from each other. Therefore, if the corner A_1 is selected in the IR image, not all of the RGB corners are considered as its potential matching corners, but only those within a circular area of radius R ($R = 50$ pixels was used in the experiments) around the RGB point A'_1 with the same coordinates as A_1 (Fig. 3).

The second strategy is based on the fact that, due to the noise and imperfection of corner detectors, the positions of the detected corners are often imprecise. This imprecision is usually not greater than a few pixels. However, if two relatively close corner points from the same image are used for hypothesis generation, the parameters of the resulting hypothesis can be significantly distorted due to such imprecision in corner positions (for example, the length and inclination of a very short line vary significantly with small displacement of its end points, resulting in a significant variation of transformation parameters s and α). In order to avoid this problem and to further reduce the number of generated hypotheses, not all pairs of corners in one image are considered to generate a hypothesis, but only those pairs that are at least d pixels away from each other ($d = 50$ pixels was used in the experiments) (Fig. 3).

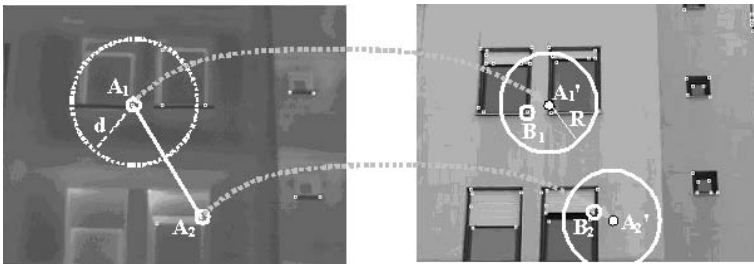


Fig. 3. Reducing the number of hypotheses

To evaluate a hypothesis, all corner points from the IR image are transformed with a hypothesized transformation and a similarity measure between the transformed points and the corner points in RGB image is computed. In multimodal images, a lot of outliers (i.e. feature points that are present only in one but not in another image) can be present. Therefore, the similarity measure has to be robust in order to reduce negative impact of these outliers. A well-known robust measure of similarity called directed partial Hausdorff distance [6,11] was used. In order to determine the directed partial Hausdorff distance between two sets of points, for each point in one set (A), the nearest point in other set (B) has to be found and the Euclidian distance between these two points has to be calculated. The partial Hausdorff distance H_k can be defined as the k -th smallest distance in the set of distances:

$$H_k(A, B) = K_{a \in A}^{th} \min_{b \in B} \text{dist}(a, b), \tag{5}$$

where K^{th} is an operator returning the k -th smallest element of the set [11]. The parameter k can also be expressed in terms of *quantile* – percentage of total number of points in the set.

Usually, the median distance is used, i.e., k is set to 50% of the number of points in the set A , assuming that there is no more than 50% of outliers in the set A (i.e., corners in the IR image). The number of outliers in the set B

(i.e. corners in the RGB image) can be greater than 50% without degrading the method performance.

In order to reduce the computational complexity of calculating Hausdorff distance for each hypothesis to be evaluated, the distance transform is used [6]. The distance transform is applied to the image containing the corners detected in the RGB image. The result is a gray-level image expressing for every pixel the distance to its nearest RGB corner. The transformation is computed only once for each image pair to be registered and can be used to determine the partial Hausdorff distance efficiently.

4 Experimental Results and Evaluation

The method was evaluated on an image database containing 40 pairs of IR and RGB images of facades taken roughly from the same viewpoint and under similar viewing angle. Despite the effort to maximize the overlap between the images of each pair, later analysis had shown that it was much worse than expected. The IR images have the resolution 320×240 and the RGB images were scaled to the same resolution.

The first evaluation was based on (subjective) visual inspection of synthetic images composed of edges found in the IR images and the corresponding RGB images. The IR edges detected by an edge detector [12] were mapped to the RGB image by the transformation found by the registration algorithm (Fig. 4).



Fig. 4. Subjective evaluation of the registration. The rightmost images are obtained by laying the edges from IR images (leftmost) over the RGB images (middle).

The registration result was subjectively evaluated either as success or failure. In total, 34 out of 40 test pairs were successfully registered by proposed method. It is worth mentioning that no parameter tuning was necessary except for three

of those image pairs. The only parameter is the percentage of point distances used for partial Hausdorff distance computation. It is by default set to 50%, but the method is successful over large variations of the parameter. For images with a large number of outliers the results improve by setting the parameter to 30%. However, six test image pairs could not be correctly registered even under extensive parameter tuning. Analysis of the detected corners in both images have shown that the reason for the failure was the large number of outliers – different corners appearing in the images (e.g., graffiti produce many corners in RGB image and none in the corresponding IR image; a tree in front of the building produces many unstable corners that overwhelm the “useful” corners corresponding to building artifacts – Fig. 5).



Fig. 5. An example of very problematic image. The corners in IR (left) and RGB (middle) images are also shown. The “registration” is obviously worthless.

Inspired by [13], we tried to gain insight into distribution of hypotheses by clustering hypotheses for “problematic” image pairs. Each hypothesis is represented by a vector of similarity transform parameters. Two approaches were taken: representation of hypotheses distribution by a histogram, and clustering hypotheses using mean-shift algorithm [14]. Unfortunately, the results were disappointing. In fact, the image pairs for which the hypotheses formed significant clusters were also successfully registered by the previously described method. On the contrary, the hypotheses generated by “problematic” image pairs don’t tend to form distinct clusters. Moreover, it turned out that some of the successfully registered image pairs didn’t generate clustered hypotheses. Therefore, we were eager to get some insight about the quality of the generated hypotheses. Also, an objective evaluation of transform quality was needed. In order to compare and evaluate the hypotheses, a rough approximation of *ground truth* information was prepared by manually marking 10 pairs of corresponding points for each image pair.

For each image pair the registration algorithm has been executed and every generated hypothesis has been recorded. Each record contained the transformation parameters and the values of partial Hausdorff distances with various quantiles (10 – 90% in the increments of 10%). Then, each hypothesis has been evaluated by measuring the average distance of *ground truth* corners from IR image, transformed by the hypothesized parameters, to the corresponding (ground truth) RGB corners. This measure will be denoted by *DGT* – *distance to the*

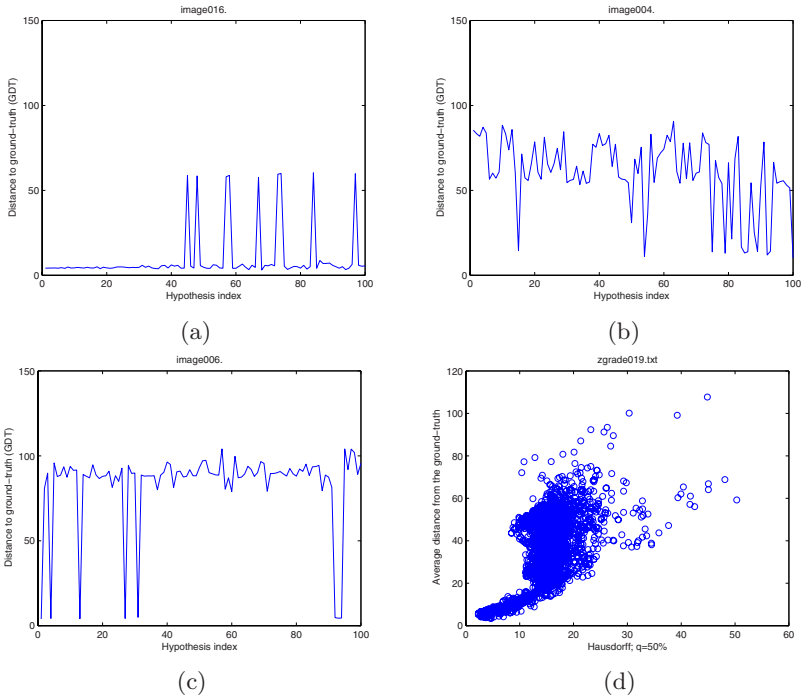


Fig. 6. (a–c) DGT graphs showing the sorted Hausdorff distances (HD) and the corresponding distances to the ground truth (DGT). (d) Scatter diagram of HD vs. DGT distances for the generated hypotheses.

ground truth. The smaller the DGT distance, the better the transformation. All hypotheses can be sorted by their Hausdorff distances (for a given quantile) and their corresponding DGT can be plotted. It is interesting that just by analyzing those graphs the success of the registration can be inferred. The successfully registered image pairs have one or more best ranked hypotheses with small DGT (Fig. 6a). If the best ranked hypothesis has a large DGT, obviously the registration procedure will fail (Fig. 6b). On the other hand, if there are many highly ranked hypotheses with small DGT, a clustering approach could filter out the possibly best ranked false hypotheses. Also, the graphs revealed some image pairs that were successfully registered due to the luck of the draw. Their graphs show that there are some generated hypotheses that are completely wrong (have large DGT) while having almost equally small Hausdorff distance (HD) as the best ranked one (Fig. 6c). The same data can be plotted as scatter diagram showing the generated hypotheses as points with coordinates (HD, DGT). Image pairs with scatter diagrams similar to the one shown in Fig. 6d are easily registered by described method because the hypotheses with small HD have small DGT (the points close to the graph origin) and the hypotheses with large DGT have large HD.

However, the diagram shown in Fig. 7a reveals that although there are many hypotheses with small HD and DGT values, many misleading hypotheses are

also generated (the points with small HD and large DGT). Those hypotheses have good Hausdorff measure although they are far away from the ground truth. Closer analysis of the corresponding image pair shows that the reason is its periodic structure (the windows from IR image can be registered with the windows in RGB image using two different horizontal translations, but only one is considered as correct – Fig. 7b). Subjective evaluation of the result considers both registrations as correct. Even humans have problems with such registration and can resolve the ambiguity only by identifying some details that can be recognized in both (multisensor) images (e.g., a broken window) – i.e., by using higher level knowledge.

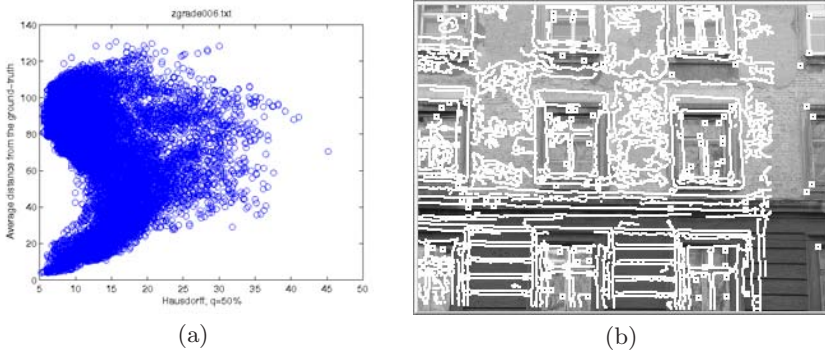


Fig. 7. (a) Scatter diagram for image pair with periodic structure. (b) The corresponding registered image.

In spite of previously discussed problems, the method has proven to be robust and reliable because in most cases (85%), the correct registration was obtained. The method is also relatively fast, although it depends on the number of detected corners, i.e. the number of generated hypotheses. Typical time required for registration of an image pair is about one second on a computer running at 1.8 GHz, with 512 MB RAM.

5 Conclusion

A simple yet effective method for infrared-visual image registration has been presented. It has been applied for aligning IR images of facades to the corresponding RGB images taken from similar viewpoint. It is intended to be a part of a system for thermal isolation inspection.

The advantage of the proposed method is that it uses small number of parameters and with same parameter setting performs well on wide range of images. The only parameter whose change influenced the results was quantile percentage. Future work would include automatic determination of quantile percentage from Euclidean distances of transformed points. We presume another major source of misregistration is insufficient characterization of corner points. Corners could be

additionally characterized e.g., using cornerity measure and corner orientation which can be calculated from matrix (2). Other usual approaches to corner characterization through corner neighborhood could fail due to different modalities of neighborhoods. The method assumes simple global similarity transformation between the images. Although this assumption does not always hold, the results show that it can be used for obtaining rough alignment of the images. Residual DGT could be made smaller using better geometric models of image transformation, e.g., refining similarity transform to affine transform or planar homography.

References

1. Brown, L.G.: A survey of image registration techniques. *ACM Computing Surveys* 24(4), 325–376 (1992)
2. Zitova, B., Flusser, J.: Image registration methods: a survey. *Image and Vision Computing* 21(11), 977–1000 (2003)
3. Krüger, W.: Robust and efficient map-to-image registration with line segments. *Machine Vision and Applications* 13(1), 38–50 (2001)
4. Coiras, E., Santamaria, J., Miravet, C.: A segment-based registration technique for visual-IR images. *Optical Engineering* 39(1), 282–289 (2000)
5. Segvic, S.: A multimodal image registration technique for structured polygonal scenes. In: *Proc of 4th Symp. on Image and Signal Processing and Analysis, Zagreb, Croatia* pp. 500–505 (2005)
6. Huttenlocher, D.P., Klauderman, G.A., Rucklidge, W.J.: Comparing images using the Hausdorff-distance. *IEEE Trans. on PAMI* 15, 850–863 (1993)
7. Irani, M., Anandan, P.: Robust multi-sensor image alignment. In: *International Conference on Computer Vision, Bombay, India*, pp. 959–966 (1998)
8. Smith, S., Brady, J.: Susan - a new approach to low level image processing. *Int. Journal of Computer Vision* 23(1), 45–78 (1997)
9. Harris, C.J., Stephens, M.: A combined corner and edge detector. In: *Proc. 4th Alvey Vision Conferences*, pp. 147–151 (1988)
10. Vincent, E., Laganier, R.: An empirical study of some feature matching strategies. In: *Proc. Conf. Vision Interface, Calgary, Canada*, pp. 139–145 (2002)
11. Mount, D.M., Netanyahu, N.S., Moigne, J.L.: Efficient algorithms for robust feature matching. *Pattern Recognition* 32, 17–28 (1998)
12. Canny, J.: A computational approach to edge detection. *IEEE Trans. PAMI* 8, 679–714 (1986)
13. Stockman, G., Kopstein, S., Benett, S.: Matching images to models for registration and object detection via clustering. *IEEE Trans. PAMI* 4(3), 229–241 (1982)
14. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Trans. PAMI* 24(5), 603–619 (2002)