



# Thermal–visible registration of human silhouettes: A similarity measure performance evaluation



Guillaume-Alexandre Bilodeau<sup>a,\*</sup>, Atousa Torabi<sup>b</sup>, Pierre-Luc St-Charles<sup>a</sup>, Dorra Riahi<sup>a</sup>

<sup>a</sup> LITIV Lab., Department of Computer and Software Engineering, École Polytechnique de Montréal, P.O. Box 6079, Station Centre-ville, Montréal, Québec H3C 3A7, Canada

<sup>b</sup> LISA, Dept. IRO, Université de Montréal, Montréal, Québec H2C 3J7, Canada

## HIGHLIGHTS

- Ten similarity measures were evaluated for thermal–visible registration.
- A new dataset was proposed for human silhouettes registration.
- Local Self-Similarity descriptor is the best choice for small objects or fragments.
- Mutual information is the best choice for large objects.

## ARTICLE INFO

### Article history:

Received 19 December 2013

Available online 1 March 2014

### Keywords:

Multispectral imagery

Similarity measures

Dense stereo matching

Thermal–visible registration

Thermal camera

Visible camera

## ABSTRACT

When dealing with the registration of information from different image sources, the *de facto* similarity measure used is Mutual Information (MI). Although MI gives good performance in many image registration applications, recent works in thermal–visible registration have shown that other similarity measures can give results that are as accurate, if not more than MI. Furthermore, some of these measures also have the advantage of being calculated independently from each image to register, which allows them to be integrated more easily in energy minimization frameworks. In this article, we investigate the accuracy of similarity measures for thermal–visible image registration of human silhouettes, including MI, Sum of Squared Differences (SSD), Normalized Cross-Correlation (NCC), Histograms of Oriented Gradients (HOG), Local Self-Similarity (LSS), Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), Census, Fast Retina Keypoint (FREAK), and Binary Robust Independent Elementary Feature (BRIEF). We tested the various similarity measures in dense stereo matching tasks over 25,000 windows to have statistically significant results. To do so, we created a new dataset in which one to five humans are walking in a scene in various depth planes. Results show that even if MI is a very strong performer, particularly for large regions of interest (ROI), LSS gives better accuracies when ROI are small or segmented into small fragments because of its ability to capture shape. The other tested similarity measures did not give consistently accurate results.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

In the recent years, there has been a growing interest in visual surveillance using multimodal sensors in both civilian and military applications. One of the fundamental issues associated with thermal–visible imagery is the matching and registration of pairs of images captured by the two different types of sensors. Unlike visible sensors that capture reflected light, IR sensors capture thermal

radiations reflected and emitted by an object in a scene. Although not very well documented, most similarity measures used for registering visible images are not applicable for thermal–visible image registration because of the differences in imaging characteristics of thermal and visible cameras. Furthermore, scene content at room temperature cannot be registered because it does not convey any textural information. For this reason, most works in thermal–visible imagery focus on hot bodies, like for example, people.

For the registration of people, two families of methods exist. First, there are the sparse registration methods. In these methods, people are considered as planar objects that can be registered by just matching some points. Most of these methods are based on matching contour points. Using edges is one of the most popular

\* Corresponding author.

E-mail addresses: [gabilodeau@polymtl.ca](mailto:gabilodeau@polymtl.ca) (G.-A. Bilodeau), [torabi@iro.umontreal.ca](mailto:torabi@iro.umontreal.ca) (A. Torabi), [pierre-luc.st-charles@polymtl.ca](mailto:pierre-luc.st-charles@polymtl.ca) (P.-L. St-Charles), [dorra.riahi@polymtl.ca](mailto:dorra.riahi@polymtl.ca) (D. Riahi).

method as their magnitudes and orientations may match between infrared and visible for some object boundaries [1–3]. Raw edges alone are not very reliable, so they may be considered as connected groups for correspondence [1,3]. Other methods use polygonal approximation of people and matches them using their vertices [4,5]. Although these methods allow fast registration of people, the resulting transformation does not allow to capture fine depth details, like the position of the arms or legs relative to the body.

For accounting for depth details, the second family is based on dense correspondences between the visible and the thermal human silhouettes. Typically, in that case, registration is performed at every pixel on the human body by comparing the pixels inside a window using a similarity measure. We define as a similarity measure any function that returns a value that indicates a level of similarity. For example, Mutual Information (MI) is one of the most popular similarity measure [6–9], but recently, Local Self-Similarity (LSS), a local image descriptor, was also proposed for this purpose. A matching window is described as a dense collection of LSS descriptors, and the similarity of two windows is the Euclidean distance between the respective collections of descriptors [10,11].

The proposal of new local image descriptors (LID) is a very active field in computer vision. These LIDs are typically proposed to allow discriminative matches between regions in visible images. For each of these new LIDs, the question is: can they be used to successfully register people in visible and thermal images? This question is worth answering for two reasons: (1) MI does not provide accurate registration all the time [6–8] and (2) sometime, surprisingly, other LIDs like Local Self-Similarity (LSS) have been shown to outperform MI [10,11]. In this article, our goal is to study the applicability of various LIDs or other similarity measures to the problem of registering people (or any other bodies not at room temperature) in visible and thermal imagery. To test the viability of various similarity measures, we use them in the context of typical windows-based matching, where the potential measures or LIDs are applied over all the pixels in a window to find correspondences.

We began some work in that regard in [12]. However, the experiments were more limited (around 300 test windows) and less similarity measures were compared. In this study, we experiment on more similarity measures and we test them on over more than 25,000 windows to have statistically significant results for human silhouette registration. We compare the similarity measures both for winner takes all (WTA) sliding window matching and disparity voting (DV).

In Section 2, we present the similarity measures we tested. Section 3 describes the details of our camera setup, dataset, test procedure, and evaluation criteria. In Section 4, we present and discuss our experimental results. Finally, we conclude the paper with a general discussion in Section 5.

## 2. Tested similarity measures

We tested three broad categories of similarity measures:

1. Similarity measures that are calculated across pixels of the two windows, namely Mutual Information (MI), Sum of Squared Differences (SSD), and Normalized Cross-Correlation (NCC);
2. Traditional LIDs that model data as distributions, namely Histograms of Oriented Gradients (HOG), Local Self-Similarity (LSS), Scale Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF). In this case, they are applied as dense collection of features compared with a distance to be used as similarity measure;
3. LIDs based on binary comparisons of pixels, namely Census, Fast RETina Keypoint (FREAK), Binary Robust Independent Elementary Feature (BRIEF). In this case they are also applied as dense collections.

These similarity measures represent only a subset of possible measures, as any LIDs could be formulated as similarity measures. However, they represent a good sample of measures as they cover the main categories of measures and LIDs and they all show good discriminative power when comparing regions in visible images.

NCC is a classic similarity measure that consists in a pixel-wise cross-correlation of two image regions normalized by the overall intensity difference [13]. It is defined for two windows on a pair of images as

$$NCC(W_1, W_2) = \frac{\sum_{x,y} (W_1(x,y) - \bar{W}_1) * (W_2(x,y) - \bar{W}_2)}{\sqrt{\sum_{x,y} (W_1(x,y) - \bar{W}_1)^2 * \sum_{x,y} (W_2(x,y) - \bar{W}_2)^2}}, \quad (1)$$

where  $W_1$  and  $W_2$  are two matching windows on a pair of thermal and visible images, and  $\bar{W}_1$  and  $\bar{W}_2$  are the mean pixel intensities in the windows. This measure relies basically on similar intensity patterns. This is similar for SSD, which is defined as

$$SSD(W_1, W_2) = \sum_{x,y} (W_1(x,y) - W_2(x,y))^2. \quad (2)$$

MI computes the statistical co-occurrence of pixel-wise image patterns inside a window on pair of images using

$$MI(W_1, W_2) = \sum_{X \in W_1} \sum_{Y \in W_2} P(X, Y) \log \frac{P(X, Y)}{P(X)P(Y)}, \quad (3)$$

where  $P(X, Y)$ , is the joint probability mass function of intensities and  $P(X)$  and  $P(Y)$  are the marginal probability functions.  $P(X, Y)$  is calculated by creating a two-dimensional histogram that records the number of co-occurrences of thermal and visible intensity values in  $W_1$  and  $W_2$ . The probabilities are then obtained by normalizing the histogram by the sum of the joint histogram entries. The marginal probabilities  $P(X)$  and  $P(Y)$  are then obtained by summing  $P(X, Y)$  over the grayscale or thermal intensities. MI relies on the co-occurrence of patterns that do not need to be similar. It can match a uniform region with a textured region. This is why it is successful with multimodal imagery.

To use LSS [14], SIFT [15], and SURF [16] as similarity measures, each of these feature descriptors are calculated densely for all the pixels in a window. For a window of say  $10 \times 10$ , this gives a collection of 100 descriptors. The similarity measure  $SM$  of two windows  $W_1$  and  $W_2$  is given by

$$SM(W_1, W_2) = \sqrt{\sum_{x,y} (f_l(x,y) - f_r(x,y))^2}, \quad (4)$$

where  $f_l$  and  $f_r$  are feature descriptors in two matching windows on a pair of thermal and visible images. SIFT and SURF are gradient-based, while LSS is a local shape descriptor based on the comparison of the self-similarity of a central patch with neighboring patches. As for HOG [17], since it is window-based, we apply it directly on the whole windows. The HOG descriptors are then compared with the Euclidean distance with Eq. (4).

Census [18], BRIEF [19] and FREAK [20] are applied the same way as SIFT. They all consist of binary intensity comparisons within a window using different pre-determined patterns. In this case, the binary descriptors of two windows are compared using the Hamming distance to obtain a similarity measure.

## 3. Experimental method

We tested the 10 chosen similarity measures on more than 25,000 windows using both a WTA and DV procedure. In this section, we describe our test methodology.

Download English Version:

<https://daneshyari.com/en/article/1784455>

Download Persian Version:

<https://daneshyari.com/article/1784455>

[Daneshyari.com](https://daneshyari.com)