MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Fast Image Registration via Joint Gradient Maximization: Application to Multi-Modal Data

Xue Mei, Fatih Porikli

TR2006-109 September 2006

Abstract

We present a computationally inexpensive method for multi-modal image registration. Our approach employs a joint gradient similarity function that is applied only to a set high spatial gradient pixels. We obtain motion parameters by maximizing the similarity function by gradient ascent method, which secures a fast convergence. We apply our technique to the task of affine model based registration of 2D images which undergo large rigit motion, and show promising results.

SPIE Conference Electro-Optical and Infrared Systems

Fast Image Registration via Joint Gradient Maximization: Application to Multi-Modal Data

Xue Mei[†] Fatih Porikli[‡]

[†]University of Maryland, College Park, MD, USA
[‡]Mitsubishi Electric Research Labs, Cambridge, MA, USA

ABSTRACT

We present a computationally inexpensive method for multi-modal image registration. Our approach employs a joint gradient similarity function that is applied only to a set high spatial gradient pixels. We obtain motion parameters by maximizing the similarity function by gradient ascent method, which secures a fast convergence. We apply our technique to the task of affine model based registration of 2D images which undergo large rigid motion, and show promising results.

Keywords: multi-modality image registration, similarity function, image alignment.

1. INTRODUCTION

Image registration amounts to establishing a common frame of reference for a set of images of the same scene taken at different time, from different views, or by different sensors. It plays a vital role in many computer vision applications such as video tracking, medical imaging, remote sensing, super-resolution and data fusion. Several comprehensive surveys[1,2] on image registration have been published to cover the progress achieved in this rich area.

In general, the registration methods can be classified into two categories[3]: direct method and feature-based method. Direct methods use pixel-to-pixel matching and minimize a measure of image similarity to find a parametric transformation between two images. Often, hierarchical approaches are adapted to improve the convergence properties. Feature-based methods first extract distinctive features from each image. Then, they match features between the set of given images to establish the correspondence and warp images according to parametric transformations estimated from those correspondences. Unlike direct methods, feature-based registration do not require initialization and able to handle large motion and viewpoint changes between the images. To find distinctive features in the image which are invariant to illumination, scale and rotation remains a chanlleging task. Brown and Lowe[4] propose to use SIFT features to register the images which is insensitive to the ordering, orientation, scale and illumination of the images and remove the 'outlier' image which doesn't have any overlapping area with the other images.

Due to the different characteristics of imaging sensors, the relationship between the intensities of corresponding pixels in multi-modality images is usually complex and unknown. Conventional intensity based feature extraction fail in case of multi-modality images. The features appear in one image might not be present at all in the other image. For example, a textured painting may appear to be a homogenous in thermal IR image in case the paints have the same radiant emittance properties. Mutual information based registration[5,6] which is inspired from information theory is considered among the state-of-the-art registration methods for multi-modality images[7] . The assumption that the intensities between corresponding pixels are similar is no longer hold, while the distribution of the intensities of matched pixels should be maximally dependent. Instead of using intensities, geometric properties such as contours and corners are also extracted to estimate the relevant transformation parameters to align the images[8,9] . To improve the convergence properties, Irani and Anandan[10] applied global estimation on the common information. Keller and Averbuch[11] proposed to register the images by iteratively minimizing the orientation distance of high intensity gradient pixels using second and third order spatial gradients. However, most existing methods assume the displacement between multi-modal images to be

For further information send correspondence to Fatih Porikli, fatih@merl.com

small, the features are highly correlated, and they are computationally prohibitive to integrate within a real-time framework.

In this paper, we present a joint, spatial gradient maximization based approach that does not require a deliberate initialization stage. Image points with strong local gradients are assumed to denote depth discontinuities, which indicate mutually observable differences in different type of images i.e. color and thermal IR. Instead of depending only on the intensity values, we take advantage of the spatial gradient responses. To find the optimum solution, we formulate the joint maximization as a gradient ascent algorithm. Our method can handle large displacements between the images and the maximization framework secures a fast convergence.

2. REGISTRATION

2.1. Maximization Function

Traditional gradient methods estimate the motion parameters p by minimizing the intensity differences between the input image $I_1(x, y)$ and warped image $I_2(x, y, p)$. The p estimation function is described as

$$p^* = arg \min_p \sum_{(x_i,y_i)\in S} (I_1(x_i, y_i, p) - I_2(x_i, y_i))^2 \tag{1}$$

where (x_i, y_i, p) is the coordinate of the corresponding pixel in the first image after we warp the first image by the motion parameters p and S is the set of coordinates of pixels that are common to both images. Gradient descent is used to find the motion parameters by solving a least square linear equation.

Alternatively, we register images by maximizing a joint spatial gradient function. Suppose edge pixels have magnitude of 1 while non-edge pixels have 0. The total energy in the sum of the correctly registered images will be much higher than the total energy obtained for incorrect registrations. Using maximization of joint gradients instead of minimization of their intensity distance provides more accurate performance since image features do not necessarily coexist in both images. In other words, our underlying assumption is not that the feature (intensity or gradient) values remain same after motion compensation between the images as in the minimization. By maximizing the joint gradient function, we obtain the parametric motion transformation parameters p. We define the motion parameters p as

$$p^* = arg \max_p \sum_{(x_i,y_i)\in S} (E_1(x_i, y_i, p) + E_2(x_i, y_i))^2 \tag{2}$$

where E_1 and E_2 represent the edge (or energy) images of I_1 and I_2 which are generated by applying Canny edge detector[12] (or by computing gradient magnitudes). Edges are the locations where image has strong gradients we consider to have depth discontinuities and high information values. Applying maximization to the a small set of salient edge pixels makes our method faster and robust.

2.2. Gradient Ascent Motion Estimation

Let the joint gradient function be $F(p) = \sum_{(x_i,y_i)\in S}(E_1(x_i, y_i, p) + E_2(x_i, y_i))^2$. $F(p)$ is nonlinear with respect to p and the motion parameters are computed using Newton's method. The iterative formulation is given by

$$p_{n+1} = p_n - (\nabla F(p))^{-1} F(p) \tag{3}$$

where p_n is the motion parameters after nth iteration and $\nabla F(p)$ is the derivative of $F(p)$ with respect to p. After a few manipulations, the iteration function for p can be computed as

$$p_{n+1} = p_n - (H^T H)^{-1} H^T E_n \tag{4}$$

where

$$H_{j,i} = \frac{\partial E_1(x_i, y_i)}{\partial p_j} \qquad \text{and} \qquad E_{n,i} = E_1(x_i, y_i, p_n) + E_2(x_i, y_i).$$

2.3. Motion Models

When the scene is approximated by a single plane, or the distance of the viewpoints between the two cameras is small relative to their distance from the scene, the motion between the two images can be modeled in terms of a single 2D transformation parameters. Hartley and Zisserman's book[13] gives a detailed description of the 2D parametric transformations. Let $\vec{p} = (p_1, p_2, \cdots, p_m)$ is the unknown parameter vector. For an affine transformation, the model is given by

$$x_i^2 = p_1 x_i^1 + p_2 y_i^1 + p_3, \qquad y_i^2 = p_4 x_i^1 + p_5 y_i^1 + p_6 \tag{5}$$

where x_i^1, y_i^1 and x_i^2, y_i^2 are the pixel coordinates before and after transformation, respectively. In this case the parameter vector becomes $\vec{p} = (p_1, p_2, p_3, p_4, p_5, p_6)^T$. The projective model is given by

$$x_i^2 = \frac{p_1 x_i^1 + p_2 y_i^1 + p_3}{p_7 x_i^1 + p_8 y_i^1 + 1}, \qquad y_i^2 = \frac{p_4 x_i^1 + p_5 y_i^1 + p_6}{p_7 x_i^1 + p_8 y_i^1 + 1} \tag{6}$$

where the parameter vector is $\vec{p} = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8)^T$.

2.4. Iterative Solution

The algorithm flow using iterative solution of gradient methods is given as follows:

p_0 : the initial estimation of the motion parameters, and

p_n : the final estimation after n iterations.

Preprocessing: Edge (gradient) images E_1 and E_2 are generated from the input images I_1 and I_2 by applying Canny edge detector (Sobel filter). The kth ($1 \leq k \leq n$) iteration is then as follows

1. Image E_1 is wrapped on E_2 using the current estimated transformation parameters p_k and the transformed image is stored in $\widehat{E_1}$.
2. $\widehat{E_1}$ and E_2 are used as input images as described in section 2.2.
3. The parameters are updated using equation 4.
4. Iterations are performed until it reaches the maximum iterations N_{max} which is set by the user, or the difference of parameters between consecutive iterations is smaller than a predetermined threshold.

2.5. Multiscale Scheme

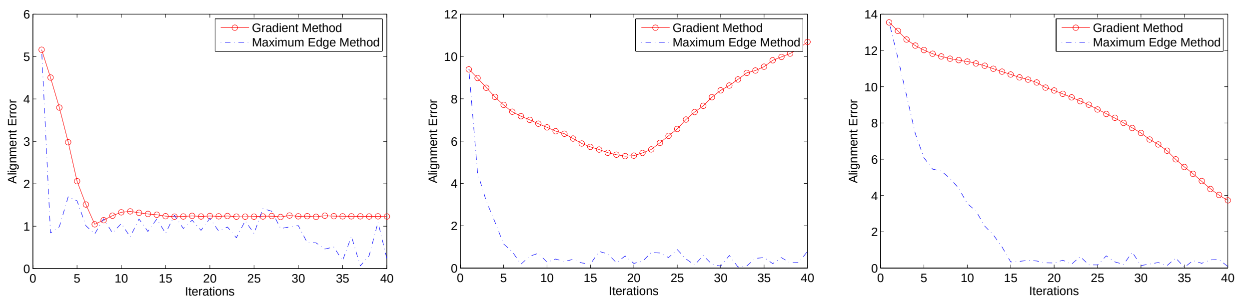
To improve the robustness and fast the convergence, the iterative solution is embedded in a coarse-to-fine hierarchical formulation. A spatial Gaussian pyramid is constructed for each image. Each pyramid level is constructed by applying a Gaussian low-pass filter to the previous level, followed by sub-sampling the image by a factor of 2. The estimation is performed on each level and goes forward to the finer level using the estimated parameters as the initial guess. We seek for initial non-zero values only for the translational parameters p_3 and p_6, leaving all the other parameters to be zeros. This is set manually or by some initialization algorithms.

3. APPLICATIONS AND RESULTS

The proposed approach was implemented and applied with an affine transformation model to image pairs acquired from visible and thermal infrared cameras. The experimental results show our algorithm achieve accurate results on both visible image registration and multi-modality image registration. Fig. 1 shows registration result of two images obtained by sensors mounted on a helicopter. There are several moving vehicles in the image which might cause the algorithm to fail. Our algorithm proves to be robust against the outliers. Fig. 3 shows the composite display of the registration result from visible and infrared images. The intensity values between these two images are very different and traditional gradient method based on invariant of intensity values fails to work. Fig. 2 and Fig. 4 show the results of translation and rotation estimation in terms of alignment error vs. the number of iterations. The alignment error is defined as the L_2 norm of the difference between ground

(a) (b) (c)

Figure 1. Registration result of two aerial images using the affine motion model. In the two images, there are several moving vehicles in the image. The registration result shows the robustness of our proposed algorithm to outliers. (a)(b) are the image pair to be registered. (c) is the registered image.



(a) (b) (c)

Figure 2. Rotation and translation estimation results for the aerial images. The image was rotated and translated by $(a)(\theta = 0^{\circ}, \delta x = 3, \delta y = 4)$ $(b)(\theta = 5^{\circ}, \delta x = 6, \delta y = 7)$ $(c)(\theta = 8^{\circ}, \delta x = 6, \delta y = 12)$ from the correct registration result. In every situation, our algorithm outperforms gradient method. Our approach converges much faster than the gradient method and more robust to the large motion.

truth transformation parameters and current estimated transformation parameters. The initial misalignment was achieved by translating and rotating the second image by a certain amount from the correct registration results. In Fig. 2, our algorithm converges significantly faster than the gradient method. The superiority becomes more apparent when the misalignment becomes larger. In Fig. 2(b), gradient method drifts away when it fails to converge to the correct transformation parameters around iteration 20. In Fig. 4, our algorithm achieves correct registration results while the gradient method fails to do so. In Fig. 4(b), the error of our method jumps up at the beginning and then drops down to converge the correct transformation parameters. It proves robustness of our method which avoids being stuck in local minimum. We observed that in most cases the gradient method drifts away from the correct registration parameters.

4. CONCLUSIONS AND FUTURE WORK

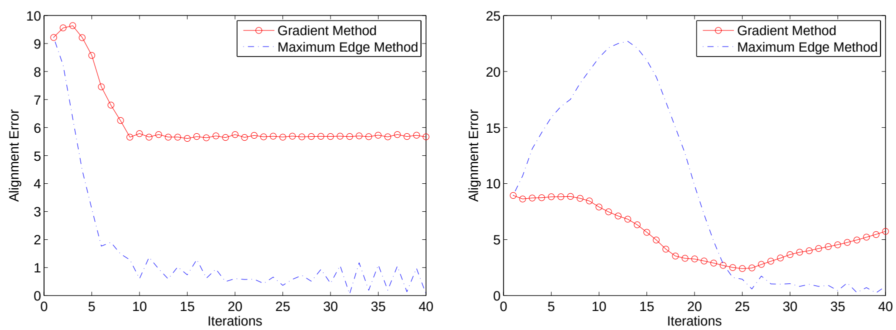
We present a fast image registration algorithm which improves the performance of the gradient based registration. Our algorithm uses a joint gradient function. Our experiments show it has superior convergence property over the conventional gradient based registration methods and it is more robust to outliers.

REFERENCES

1. L. Brown, "A survey of image registration techniques," *ACM Computing Surveys* **24**, pp. 325–376, 1992.
2. B. Zitova and J. Flusser, "Image registration methods: A survey," *Image and Vision Computing* **21**, pp. 977–1000, 2003.
3. R. Szeliski, "Image alignment and stitching: a tutorial," *Microsoft Research MSR-TR-2004-92* , 2004.

(a) (b) (c)

Figure 3. Multi-sensor Alignment. (a) Original visual image. (b) Original IR image. (c) Composite display after alignment. Original IR image was rotated and translated before alignment. Though the apparent visual difference and very different intensity values between the visible and IR images, our approach achieves good results on the experiments.



(a) (b)

Figure 4. Rotation and translation estimation results for the multi-modality image registration. The IR image was rotated and translated by $(a)(\theta = 0°, \delta x = 6, \delta y = 7)$ $(b)(\theta = 2°, \delta x = 4, \delta y = 8)$ from the correct registration result. In both cases, our algorithm achieves correct registration results, while gradient method drifts away from the correct results.

4. M. Brown and D. G. Lowe, "Recognising panoramas," *IEEE International Conference on Computer Vision* , pp. 1218–1225, 2003.
5. A. Collignon, F. Maes, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multimodality image registration using information theory," *Information Processing in Medical Imaging* , pp. 263–274, 1995.
6. P. Viola and W. M. W. III, "Alignment by maximization of mutual information," *International Conference on Computer Vision* , pp. 16–23, 1995.
7. J. Pluim, J. Maintz, and M. Viergever, "Mutual information based registration of medical images: a survey," *IEEE Trans. on Medical Imaging* **8**, pp. 986–1004, 2003.
8. V. Govindu and C. Shekhar, "Alignment using distributions of local geometric properties," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **21**, pp. 1031–1043, 1999.
9. A. Guziec, X. Pennec, and N. Ayache, "Medical image registration using geometric hashing," *IEEE Computational Science and Eng.* **4**, pp. 29–41, 1997.
10. M. Irani and P. Anandan, "Robust multi-sensor image alignment," *IEEE International Conference on Computer Vision* , pp. 959–966, 1998.
11. Y. Keller and A. Averbuch, "Multisensor image registration via implicit similarity," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **28**, pp. 794–801, 2006.
12. J. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **8**, pp. 679–698, 1986.
13. R. Hartley and A. Zisserman, *Multiple View Geometry*, Cambridge University Press, Cambridge, UK, 2004.