Final Technical Report
TNW2008-03

Research Project Agreement No. 61-2307

Optimization-Based Methods for Road Image Registration

George Turkiyyah and Tyler Sprague

Department of Civil and Environmental Engineering
University of Washington
Seattle, WA 98195

A report prepared for

Transportation Northwest (TransNow)
University of Washington
135 More Hall, Box 352700
Seattle, Washington 98195-2700

February 2008

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.	
TNW2008-03			
4. TITLE AND SUBTITLE Optimization-Based Methods for Road Image Registration		5.REPORT DATE February 2008 6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) George Turkiyyah, Tyler Sprague		8. PERFORMING ORGANIZATION REPORT NO. TNW2008-03	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Transportation Northwest Regional Center X (TransNow) Box 352700, 129 More Hall University of Washington Seattle, WA 98195-2700		10. WORK UNIT NO. 11. CONTRACT OR GRANT NO. DTRS99-G-0010	
12. SPONSORING AGENCY NAME AND ADDRESS United States Department of Transportation Office of the Secretary of Transportation 400 Seventh St. S.W. Washington, D.C. 20590		13. TYPE OF REPORT AND PERIOD COVERED Final Research Report 14. SPONSORING AGENCY CODE	
15. SUPPLEMENTARY NOTES			

ABSTRACT

A number of transportation agencies are now relying on direct imaging for monitoring and cataloguing the state of their roadway systems. Images provide objective information to characterize the pavement as well as roadside hardware. The tasks of processing, interpreting, and assessing the condition of the pavement from image data sets poses formidable challenges however. Not only are the data sets extremely large but they are not taken under ideal conditions neither in geometric location and alignment, nor in photometric consistency due to a number of factors including the inaccuracy of dead reckoning as well as the dynamics of the collection vehicles. The resulting images provide, in essence, a set of independent overlapping snapshots of regions of the pavement that must be merged together to produce the continuous mosaic that is needed for automated analysis and interpretation. This requires the registration of these images. Registration the problem of aligning features in one image/view to corresponding features in another image/view of the same object. Registration brings independent images into the same reference frame. To be effective, this process must be automated requiring minimal or no user intervention. In this work, we describe a non-parametric optimization procedure that is capable of producing general nonlinear deformations to register multiple images so they can be merged, compared and analyzed.

17. KEY WORDS		18. DISTRIBUTION STATEMENT	
19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None	13	

DISCLAIMER

Optimization-Based Methods for Road Image Registration

Tyler Sprague and George Turkiyyah
University of Washington
December 2007

Abstract

A number of transportation agencies are now relying on direct imaging for monitoring and cataloguing the state of their roadway systems. Images provide objective information to characterize the pavement as well as roadside hardware. The tasks of processing, interpreting, and assessing the condition of the pavement from image data sets poses formidable challenges however. Not only are the data sets extremely large but they are not taken under ideal conditions neither in geometric location and alignment, nor in photometric consistency due to a number of factors including the inaccuracy of dead reckoning as well as the dynamics of the collection vehicles.

The resulting images provide, in essence, a set of independent overlapping snapshots of regions of the pavement that must first be merged together to produce the continuous mosaic that is needed for automated analysis and interpretation. This requires the registration of these images. Registration the problem of aligning features in one image/view to corresponding features in another image/view of the same object. Registration brings independent images into the same reference frame. To be effective, this process must be automated requiring minimal or no user intervention. In this work, we describe a non-parametric optimization procedure that is capable of producing general nonlinear deformations to register multiple images so they can be merged, compared and analyzed.

1 Introduction

A number of transportation agencies are now relying on direct imaging for monitoring and cataloguing the state of their roadway systems. Images provide objective information to characterize the pavement as well as roadside hardware. As the nation's infrastructure ages, ensuring safety and ride quality have become more prominent items on the agenda of many agencies, and detailed imaging provides the raw data for addressing them.

While highly engineered structures such as bridges and elevated freeways are limited in number and as such can be reliably and realistically evaluated with individualized instrumentation and custom snapshots, pavement systems require a different condition monitoring strategy. In order to provide comprehensive coverage of of the roadway system, video imaging is unavoidable. Given the amount of traffic in urban areas, the only practical and cost-effective way to generate this imaging today is by mounting video cameras and related instruments on a van that travels the highway system at normal speeds alongside with the rest of the traffic. The data collected can then form the starting point for meaningful evaluation and condition assessment.

An example of the data collection technology that is deployed today comes from Pathway Services Inc. Figure 1 shows an instrumented van that carries a number of cameras mounted on structural frames on its front and back sides. Different configurations of the collection system exist with a differing number of video cameras, differing orientations and relative locations of these cameras, etc. Generally however, a couple of the cameras are pointing downwards and a couple are pointing forward. The position, orientation, inclination of the van, all essential data for later processing, are continuously recorded. The van's position is generated by dead-reckoning.

Figure 1: Image Data Collection Van

Figure 2 shows a typical screenshot of the information being continuously recorded. The top portion of the screen shows two video frames collected by right and left forward-pointing front cameras. The bottom portion shows the two frames collected by right and left downward-pointing cameras. To reduce the amount of data storage, the pavement condition images are recorded at about 10 frames per second (fps). The cameras are mounted at about 6 ft from the ground. The resulting video stream consists of sequence of frames each covering 5-6 ft of roadway surface in the direction of travel. Notice that the image pairs (both front- and down-pointing images) are not seamless. The features they record may overlap. The intensity level of a given pair are different. Similar effects are also present in successive snapshots of the same camera.

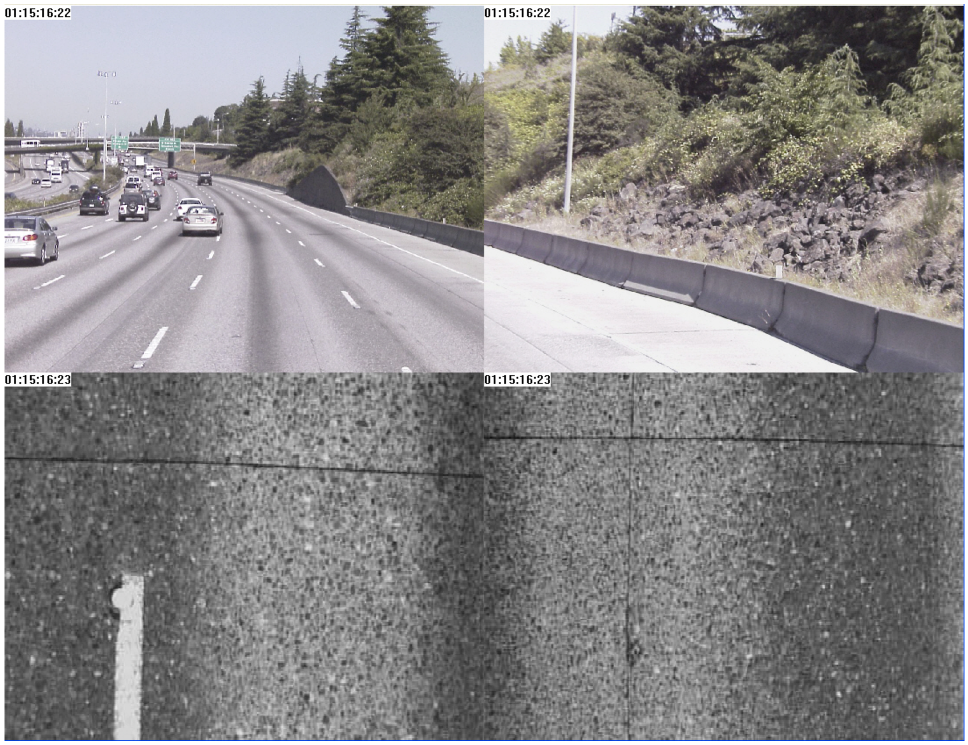

Figure 2: Typical 4-camera video frames captured by front and back mounted on the van

The tasks of processing, interpreting, and assessing the condition of the pavement from these data sets poses formidable challenges. Not only are the data sets extremely large but they are not taken under ideal conditions neither in geometric location and alignment, nor in photometric consistency. The geometric misalignment is due to a number of factors including the inaccuracy of dead reckoning as well as the dynamics of the van. The van has its own dynamic behavior as it travels at highway speeds, and navigates turns and slopes in the road. The van dynamics change the position and orientation of the cameras with respect to the road surface resulting in overlaps and misalignments between successive frames. Photometric inconsistencies coming from the lenses of the cameras, further distort the images and complicate the merging process.

2 Problem Statement and Objectives

The resulting images provide, in essence, a set of independent overlapping snapshots of regions of the pavement that must first be merged together to produce the continuous mosaic that is needed for automated analysis and interpretation. In order to be effective, the task of processing these data must be somewhat automated to allow the initial extraction of meaningful deterioration and damage information from image data. Quantities such as the number of cracks, their lengths, their widths etc. cannot be reliably quantified in an automated fashion unless the image data is first processed into a seamless stream. This is one of the main challenges in the initial processing of the data. We address this problem by devising methods for the automated registration and merging of overlapping images.

Image registration is a challenging problems in image interpretation. Image registration can be defined as the problem of aligning points and features in one image/view to corresponding points and features in another image/view of the same object. It involves deforming an image to best fit another one. Image registration is essential whenever images obtained at different times, devices, lighting conditions, perspectives, etc, need to be compared and integrated in some fashion. In many problems, as is the case here, specially-planted markers to help with the registration do not exist. Image data and perhaps some a-priori information about the content of the images are the only information available.

The image registration problem may be formulated as an optimization problem where the objective is to minimize a distance metric between the two images. A number of different formulations have been used in the past define the unknowns in the problem. The approach we use here is to define the unknown registration as a discretization of a non-rigid displacement field that distorts one image into another. The formulation is generally ill-posed and needs to be regularized, and constraints from a-priori problem information are also included.

Techniques from finite element methods in elasticity may be used in the discretization, and the resulting problem consists of a nonlinear algebraic optimization problem that may be solved by Newton's method. The function to be minimized has two parts. One of them computes internal forces defined for the unknown displacement field, while the other one is responsible for external forces computed from image data. The internal forces regularize the solution by keeping the displacement field smooth, while the external forces are defined to obtain the desired registration. Similar ideas have been used in medical image registration problems, and made practical and affordable by the massive amount of computational power now available in most offices/departments.

3 Background

The use of imaging for data collection and analysis of pavement condition is not new. As early as 1990, Ritchie [8] described the potential application of image-processing methods to automate the manual data collection and visual rating of pavement surface conditions. A number of proposals for cost-effective automated systems to capture and extract pavement-surface distress from video images have been described in a number of contexts since. Despite the great potential of these proposals, much work remains to realize the efficiency and consistency of automated condition assessment.

A number of systems were developed for quantifying distress and cracking using image processing methods. The system of Huang and Xu for instance [2] uses a multiresolution segmentation for crack detection. It operates on single images and divides the image into 8x8 cells that are classified as cracks or non-cracks and merged together into clusters when they fit template patterns. The merged clusters are then further processed into classifiable crack segments that can be joined and measured. While these techniques can be made to work well for single images, cracks and other classifiable damage often spans multiple frames (say a 15-ft longitudinal crack in the pavement that may span 3 or 4 image frames). Without registering and merging of the successive image frames into a seamless composite image, such damage patterns cannot be reliably quantified. The use of single image interpretation methods however can still play an important role for quick real-time evaluation during data collection.

There has also been attempts to use artificial neural network-based (ANN) methods as the basis of crack imaging systems to identify and classify crack types of digital pavement images [5, 6]. Tiles are used to smooth the raw images and threshold learned by the neural networks are used classify regions into crack or background. However, these systems have not yet reached sufficient robustness.

The application of imaging techniques for assessing the uniformity of newly constructed pavements has also been recently demonstrated [4]. The system was designed to be used for the detection of segregated Hot Mix Asphalt (HMA) areas and identification of these areas along a road for HMA quality assurance purposes. The image processing module uses a Gray Level Co-occurrence Matrix (GLCM), commonly used in visual texture analysis, to compute various different parameters that characterize surface texture. Measures of contrast, correlation, energy, and homogeneity are used to discriminate areas having different visual texture characteristics in the pavement surface and flag unacceptable non-uniformities.

The biggest obstacle today towards the routine use of imaging in all aspects of pavement evaluation does not come from the hardware or operational acquisition costs. Inexpensive cameras and mounting frames can now be readily purchased and deployed. Rather, the main impediments come from the fact that the process of evaluating and comparing images now requires significant time from human operators. The images captured, while covering every inch of road, are overlapping as explained above. Further, when they are taken periodically, they are never taken at the exact same locations nor with the same lighting conditions, etc. Currently human operators have to look painstakingly at the massive amounts of image data to make these assessments. They have to mentally "merge" multiple successive images to form a mosaic that can then be assessed and/or compared to existing images and known patterns.

This is time-consuming, expensive, error-prone and perhaps not the most effective way to produce useful quantitative information from the data. In practice, the use of images to assess damage and changes in distress condition is not possible without registering images to combine individual frames into large units, to allow meaningful comparisons of these composite frames to each other.

4 Methods

The main contribution of this work is the development and demonstration of an automated method for the registration and merging of successive images that are overlapping, nonlinearly distorted, and with different intensity levels without manual intervention or the explicit identification of corresponding pairs of landmarks in the images. We describe the mathematical formulation of the problem, its solution, and show illustrations on sample images.

4.1 Formulation

Given two scalar images $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$, the goal of image registration is to find the optimal geometric transformation M such that the difference between $I_1(\mathbf{x})$ and $I_2(M(\mathbf{x}))$ in their corresponding/overlap region is as small as possible. The difference between the images can be measured in a variety of ways that differ in their complexity and robustness. Distances based on identifying "landmarks" in both images and minimizing the distance between them are not convenient because they may require the manual definition of such pairs. For grayscale images the integral of the difference in intensity value is a natural and easily computable metric. $D(I_1, I_2, M) = \frac{1}{2}\int_\Omega (I_1(\mathbf{x}) - I_2(M(\mathbf{x})))^2 d\Omega$. However it is not always robust and we describe a better metric that is based on gradient information below.

In order to find the optimal transformation $M(\mathbf{x})$, we may choose the functional form of M and then use an appropriate numerical optimization scheme to find the specific values that minimize the distance metric between the two images. For simple images, it is sometimes possible to use a linear transformation model, defined as:

$$M(\mathbf{x}) = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + \begin{Bmatrix} a_5 \\ a_6 \end{Bmatrix}$$

Such a mapping produces affine transformations and except in very special circumstances is not adequate for registration. Higher order polynomial mappings can produce a larger set of potential transformations, however the fixed functional form of the parametric geometric transformation is quite constraining. In our experiments a bicubic mapping was able to produce acceptable registration between successive images on some images. Generally however, it is not possible to find an a-priori parameterizable mapping that will reliably generate registration between the image pair. As a result, we found that nonparametric registration techniques are needed for generality.

Nonparametric registration is a more general technique that requires neither the identification of landmarks nor the imposition of a pre-parameterized mapping on the optimization problem. In non-parametric registration, we seek a smooth deformation *field*, defined over the whole image, that transforms one image into the other. Mo parameters are involved in the representation of the transformation. Mathematically, non-parametric registration is more involved because the resulting numerical optimization problem is ill-posed and requires regularization. In nonparametric registration the mapping is written as:

$$M(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$$

where u is a displacement vector defined over the domain of the image. Computationally, the field $\mathbf{u}$ is discretized and the distance metric to be minimized is then defined as the L_2 norm of the difference between the two images.

$$f(\mathbf{u}) = D(I_1, I_2, \mathbf{x} + \mathbf{u}) = \frac{1}{2}||I_1 - I_{2u}||$$

To verify this formulation, we developed an initial implementation to perform the registration on synthetically generated images. The two synthetic images in the left column of Figure 3 below are shown and the goal is to register the top image (disk) to the bottom one (larger disk with a cut out). By minimizing the objective function $f(\mathbf{u})$ we are able to readily generate the x- and y-components of $\mathbf{u}$ to yield an essentially perfect match. The final x- and y- components are plotted in the right column of Fig 3.

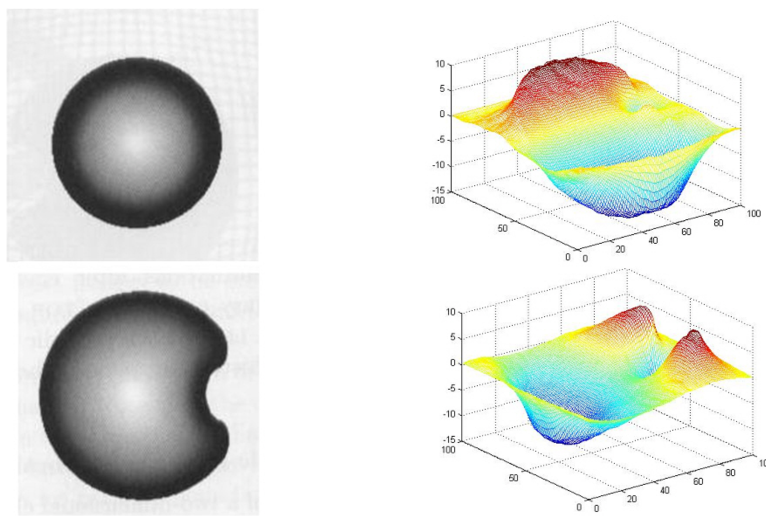

Figure 3: Nonparametric Registration applied to Synthetic Images (elastic regularization is used)

4.2 Distance Metrics

While the intensity metric described above was successfully used in the idealized synthetic example, images with noise, fewer discrete features, different levels of intensity, etc. do not produce the best registration using that metric. A more robust image similarity measure is sought which handles images with different overall intensity levels while still being suited for use in numerical optimization procedures. We investigated an alternative distance measure which is based on normalized gradients. The idea of using gradients to characterize similarity between images is based on the observation that image structure can be defined by intensity changes rather than absolute values of intensity. The resulting normalized gradient metric is more robust, easy to interpret, and computationally efficient.

Let $\mathbf{n}$ be the normalized local gradient field of the image, $\mathbf{n}(I, \mathbf{x}) = \frac{\nabla I(\mathbf{x})}{||\nabla I(\mathbf{x})||}$ when the gradient has a non-zero magnitude. Qualitatively, a good similarity is achieved when the gradients line up, i.e. when their dot product is maximized. This directly leads to a definition of a gradient-based scalar distance metric as follows:

$$f(\mathbf{u}) = D(I_1, I_2, \mathbf{u}) = -\frac{1}{2}\int_{\Omega} [\mathbf{n}(I_1, \mathbf{x}) \cdot \mathbf{n}(I_{2u}, (\mathbf{x})]^2 \, d\mathbf{x}$$

Image regions with uniform intensity in the image produce gradient directions that are very sensitive to noise and small values of the gradient. In practice, this difficulty is avoided numerically by adding a small ϵ to the computation of the gradient magnitude as suggested in [3]. Our experiments show that gradient-based similarity metrics are more robust than intensity-based ones.

4.3 Elastic Regularization and Solution

As alluded to earlier, the straight minimization of $f(\mathbf{u})$ is an ill-posed problem. Therefore, a regularization term must be added to the objective function in order to be able to solve the optimization problem. A number of regularizers to smooth the acceptable displacement fields, have been proposed in the literature [7]. We used an elastic regularizer in our experiments. The objective function to be minimized becomes:

$$f^*(\mathbf{u}) = f(\mathbf{u}) + \alpha S(\mathbf{u})$$

where S is the elastic potential of the displacement field $\mathbf{u}$, hence the name elastic registration.

The solution of the optimization problem above reduces to solving the Euler-Lagrange equations:

$$K[\mathbf{u}](\mathbf{x}) - g(\mathbf{x}, \mathbf{u}) = 0, \text{for all x}$$

The PDE operator K is related to the smoother S and can be interpreted as the stiffness matrix of an elastic medium, while g is related to the distance metric and can be interpreted as forces that deform the elastic medium. The iterative solution of the above equation produces the optimal registration.

The implementation of the elastically regularized registration starts with a mesh of discrete elements. The nodes of the mesh are chosen to be centers of pixels and artificial material properties assigned to "elements" connecting the nodes. For efficiency reasons, a slightly coarser resolution of the image is used for defining the mesh (and the deformation field). For an image or N nodes, the standard finite element displacement field is defined as:

$$\mathbf{u}(\mathbf{x}) = \sum_{i=1} N\phi_i(\mathbf{x})u_i$$

where ϕ_i is the nodal basis function and $\mathbf{u}_i$ is the displacement of node i. The stiffness matrix for each element is computed and is assembled into a global stiffness matrix K as is standard in displacement-based finite element implementations. The element stiffness matrix is formed of 2×2 k_{ij} blocks:

$$k_{ij} = \int_{\Omega} B_i^t D B_j d\Omega$$

$$B_i = \begin{bmatrix} \phi_{i,x} & 0 \\ 0 & \phi_{i,y} \\ \phi_{i,y} & \phi_{i,x} \end{bmatrix}$$

Boundary conditions must also be specified to ensure a well-posed formulation. Since our registration problem involves moving and deforming one image relative to another, the fixed boundary conditions around the image perimeter that are typically used in other applications cannot be used here. We experimented with free per-node boundary conditions as well as per-edge free displacement and compared their results.

In order to solve for the displacement field at every iteration, we must apply the g forces computed from the difference between the two images. This is the set of external fictitious driving forces that lead to registration. g is computed as:

$$g(\mathbf{x}, \mathbf{u}) = (I_1 - I_{2u}) \nabla I_{2u}(\mathbf{x})$$

4.4 Examples

An example of the performance of the optimization procedure is shown below. Consider the two overlapping frames in Figure 4. The task is to deform the (say) top frame so that its features line up with the corresponding features of the bottom frame, as the first step in condition assessment. Since we seek an automated procedure, a manual definition of landmarks, keypoints, or any pair of correspondences between the two frames is not acceptable. Notice that even for this simple pair of frames, no affine transformation of one image will produce the other one. Even the curvature of the transverse joint (same spatial object) look different in the images. The procedure has to automatically generate the nonlinear deformation that produces the intuitive registration that our eyes readily see.

As a rule, affine transformation are insufficient for registering the types of images acquired from video cameras in the setup shown in Figure 1. Nonlinear deformation is needed because of a number of effects including possible lens distortion, vehicle dynamics causing roll and pitch, etc. These effect are particularly pronounced when the "image" is a composite of multiple frames. The change in shape of the feature can also be due to increased distress over time. Nonlinear and nonparametric registration methods are needed to handle the general nature of the image streams.

When the optimization method described above is applied to the pair of images in 4 with a boundary condition on top that constraints the deformation to be uniform in the horizontal direction, we end up with the registration (and merging) shown in the left of Figure 5. A free boundary that doesn't impose such constraint produce the registration shown on the right. The difference are, as expected, slight because the driving terms of the registration

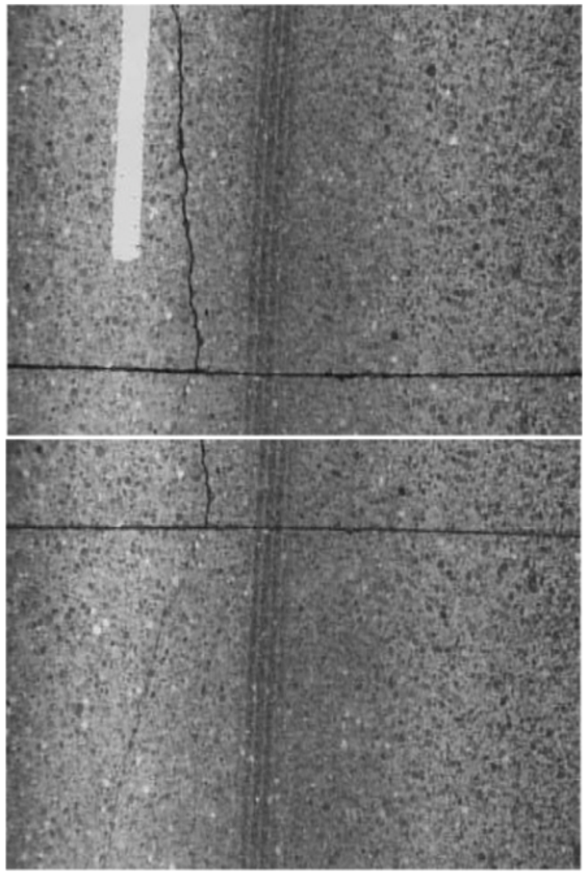

Figure 4: Two overlapping images that need to be registered and merged

come only from the overlap region which covers only the bottom part of the top image. Figure 6 shows the resulting deformation field that produced the optimal registration and match.

5 Conclusions and Future work

In this work we demonstrated a general and unified method for matching images using an optimization procedure to deform one image into another. The method requires no manual intervention and can produce general nonlinear deformation to minimize a distance metric. We experimented with a number of different distance metrics as well as regularization strategies. This registration method can be used to fuse multiple overlapping images together, to track changes between two images of the same distress, and to match a set of distresses in one image to the corresponding distress in another.

Much work remains to be done however. First, a more detailed study with larger set of images needs to be done to determine experimentally the performance and effectiveness of the algorithm under various conditions of image acquisition (say, moderate to extreme

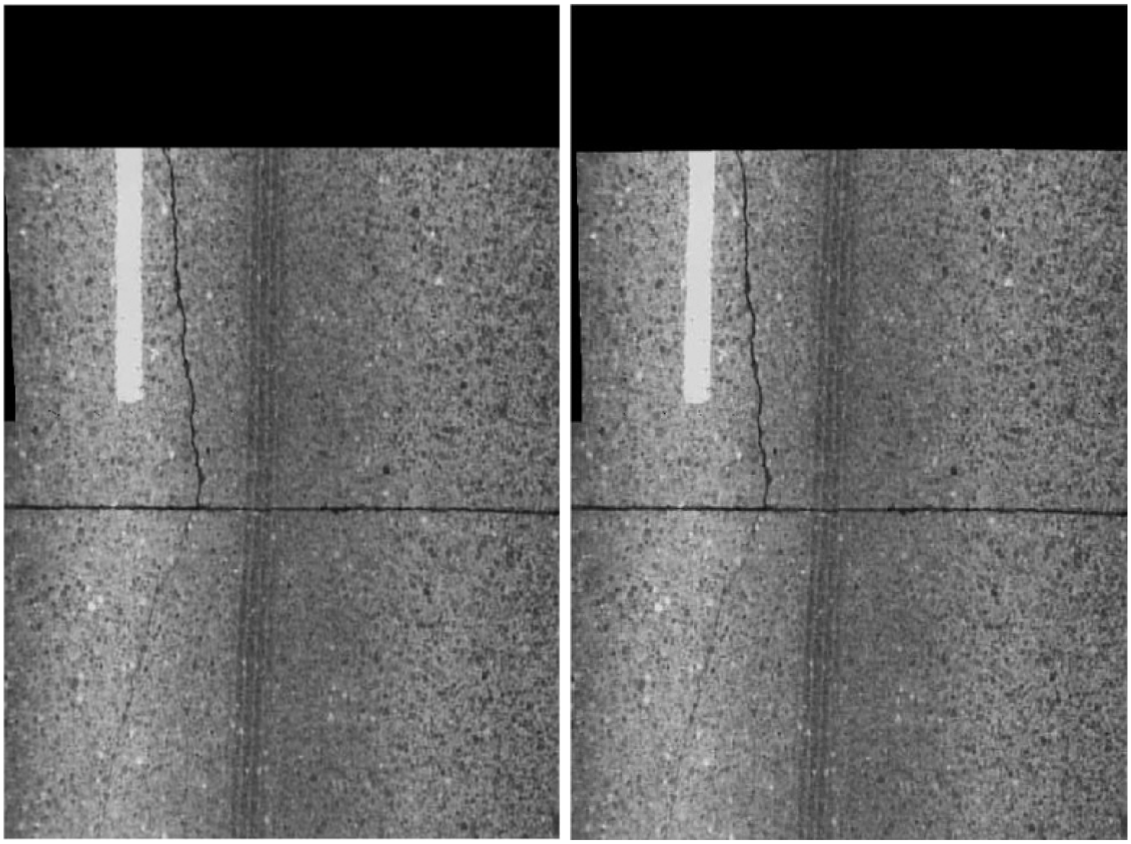

Figure 5: Optimal elastic registration

vibration). Second, images acquired as part of the routine task of driving the roadway system are not only noisy but also include a number of features that complicate inter-year registration. Consider for the example Figure 7 which shows set of overlapping video frames of the same section of road taken during two successive years (one per column). The effects of shadows, lane restriping, etc. are challenging to a raw registration process. The regions of the images where these effects exist must be first identified and not be included in the registration/matching process.

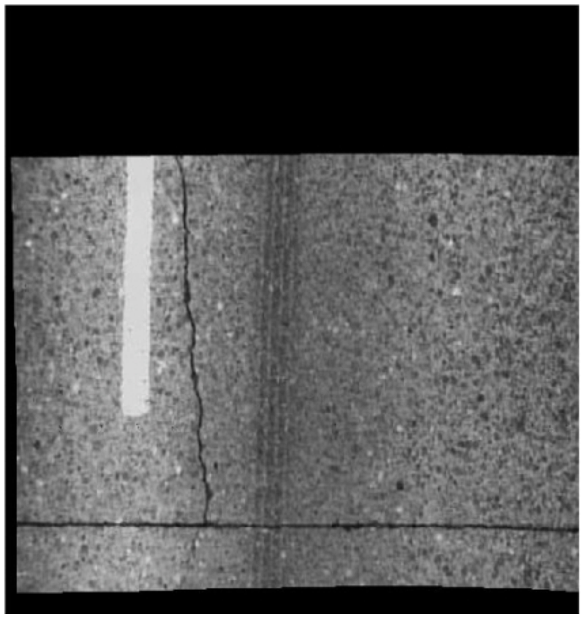

Figure 6: Nonlinear nonparametric deformation that resulted in the optimal registration

References

[1] A. Levin A. Zomet S. Peleg Y. Weiss, *Seamless image stitching in the gradient domain*, ECCV (T. Padjla J. Matas, ed.), Springer-Verlag, 2004.

[2] Y. Huang B. Xu, *Automatic inspection of pavement cracking distress*, Journal of Electronic Imaging **15** (2006), no. 1.

[3] U. Ascher E. Haber H. Huang, *On effective methods for implist piecewise smooth surface revovery*, SISC **28** (2006).

[4] E. Izeppi G. Flintsch W. Gerardo W L. Abbott, *Measuring the uniformity of hot-mix asphalt pavements with digital image technology*, TRB 86th Annual Meeting, 2007.

[5] H. Cheng J. Wang Y. Hu C. Glazier X. Shi X. Chen, *Novel approach to pavement cracking detection based on neural network*, Transportation Research Record **1764** (2001).

[6] W. Xiao X. Yan X. Zhang, *Pavement distress image automatic classification based on density-based neural network*, Lecture Notes in Computer Science, vol. 4062, Springer Verlag, 2006.

[7] J. Modersitzki, *Numerical mehods for image registration*, Oxford, 2004.

[8] S. Ritchie, *Digital imaging concepts and applications in pavement management*, Journal of Transportation Engineering **11** (1990), no. 3.

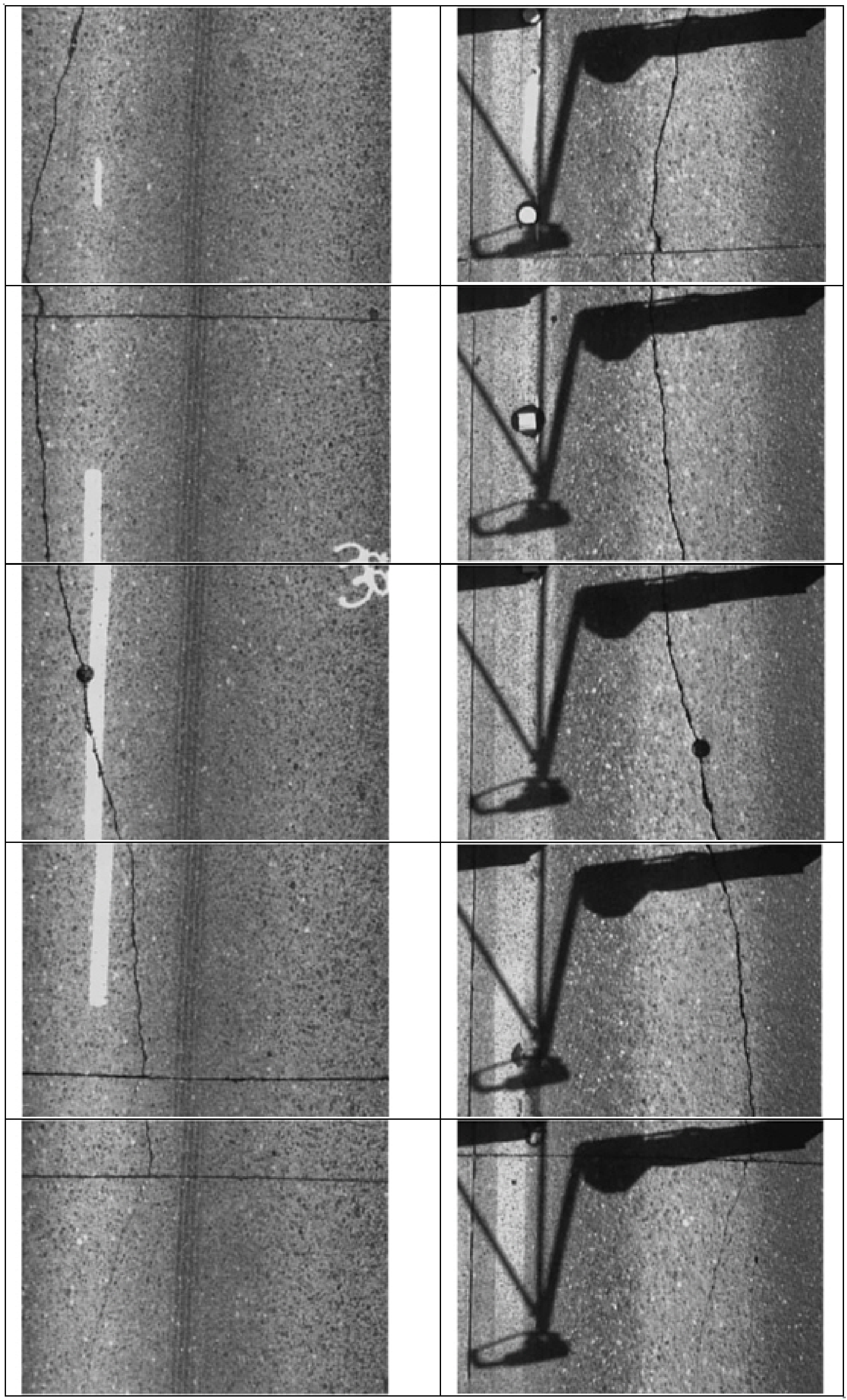

Figure 7: Difficulties: shadows, lane restriping, and defacing of features