A Mathematical Comparison of Point Detectors

M. Zuliani, C. Kenney and B. S. Manjunath
Vision Research Lab
Department of Electrical and Computer Engineering
University of California, Santa Barbara
Santa Barbara, CA 93106
Email: {zuliani, kenney, manj}@ece.ucsb.edu

***Abstract*— Selecting salient points from two or more images for computing correspondences is a fundamental problem in image analysis. Three methods originally proposed by Harris et al. in [1], by Noble et al. in [2] and by Shi et al. in [3] proved to be quite effective and robust and have been widely used by the computer vision community. The goal of this paper is to analyze these point detectors starting from the algebraic and numerical properties of the image auto-correlation matrix. To accomplish this task we will first introduce a "natural" constraint that needs to be satisfied by any point detector based on the auto-correlation matrix. Then, by casting the point detection problem in a mathematical framework based on condition theory [4], we will show that under certain hypothesis the point detectors [2], [3], [4] are equivalent modulo the choice of a specific matrix norm. The results presented in this paper will provide a novel unifying description for the most commonly used point detection algorithms.**

I. INTRODUCTION

In order to establish correspondences among a collection of images it is first necessary to identify a set of salient points in each image. Point correspondences constitute the input for more complex algorithms that aim at registering images, at reconstructing the three dimensional structure of a scene or at monitoring activities in a certain area, just to list a few. One of the paradigms to compute point correspondences that is largely used by the computer vision community is composed of two steps: initially a set of tie points is detected in each input image and then this set is fitted to a model that is supposed to describe the transformation between the images. A typical example is the estimation of the planar homography between two views of a planar scene. Crucial for the success of this approach is a high repeatability rate of the point detector (see [5]). This means that if a point $\mathbf{q}$ is detected in image $\mathcal{I}$ then, with high probability, the corresponding point $\mathbf{q}'$ will be detected in image $\mathcal{I}'$. Since the transformation that relates the images is not known a priori, a point detector should exploit only the information contained in one single image. Intensity based methods achieve this goal by associating to each point $\mathbf{q}$ a scalar value that defines its "goodness" as a tie point. This evaluation is done by performing some operations on the image intensity values in a neighborhood of the point $\mathbf{q}$. The methods proposed by Harris and Stephens [1] (and the modified version introduced by Noble [2]), by Shi and Tomasi [3] and by Kenney et al. [4] can all be considered intensity based methods.

The goal of this work is to explore in detail what is the quantitative relation between these detectors, starting from the algebraic and numerical properties of the image auto-correlation matrix. This is motivated by the fact that all the detection rules are based on the spectral structure of this matrix. Our first contribution to this analysis is to observe that any function (based on the auto-correlation matrix) used to detect tie points should satisfy a monotonicity constraint that arises naturally from the "physical" interpretation of the auto-correlation matrix. Then, after having noticed that the rationale behind the point detectors proposed by Harris, Stephens and Noble and by Shi and Tomasi entail a notion of computational stability, we shall see how these considerations can be restated from first principles using a mathematical framework based on condition theory. This approach will shed new light on the rules used to detect tie points, demonstrating that the previously mentioned approaches are equivalent modulo a suitable choice of a matrix norm.

The paper is structured as follows: section II introduces some of the notation that will be used throughout the paper, section III summarizes the condition theory framework used to relate the different point detectors, which are presented in section IV together with the nondecreasing constraint. The main results of the paper are stated, proved and discussed in section V, whereas the final conclusions can be found in section VI.

II. PRELIMINARIES

A. Geometric Transformations of Images

A single channel image $\mathcal{I}$ can be thought as a bounded scalar function defined over a compact subset of $\mathbb{R}^2$:

$$\mathcal{I} : \mathcal{D} \to [\mathcal{I}_{min}, \mathcal{I}_{max}] \subset \mathbb{R} \qquad \mathbf{q} \mapsto \mathcal{I}(\mathbf{q})$$

We are interested in detecting tie points in images related by a geometric transformation $\mathbf{T}_{\boldsymbol{\theta}} : \mathbb{R}^2 \to \mathbb{R}^2$ that depends on the parameter vector $\boldsymbol{\theta}$. These transformations can be grouped in *transformation classes* defined as:

$$\mathcal{T} \stackrel{\text{def}}{=} \{\mathbf{f} : \exists \boldsymbol{\theta} \text{ such that } \mathbf{f} = \mathbf{T}_{\boldsymbol{\theta}}\}$$

In this paper we will focus our attention on two classes of transformations:

- Translation ($\mathcal{T}_T$): $\mathbf{T}_{\boldsymbol{\theta}}(\mathbf{x}') \stackrel{\text{def}}{=} \mathbf{x}' + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$
  for which the parameter vector is $\boldsymbol{\theta} = \begin{bmatrix} t_x & t_y \end{bmatrix}^T$

- Rotation and translation ($\mathcal{T}_{RT}$):
  $\mathbf{T}_{\boldsymbol{\theta}}(\mathbf{x}') \stackrel{\mathrm{def}}{=} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} (\mathbf{x}' - \mathbf{q}') + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$
  for which the parameter vector is $\boldsymbol{\theta} = \begin{bmatrix} \phi & t_x & t_y \end{bmatrix}^T$ and $\mathbf{q}'$ is the point about which the image patch rotates.

For an example of a transformation in the class $\mathcal{T}_{RT}$ see Figure 1. Note that we did not consider transformations that involve scale, since scale requires a specific treatment. This is because the point detection methods we will be considering are intensity based methods. This means that they process the image intensities in a neighborhood centered about the point of interest and therefore such neighborhood should transform covariantly with the images. As an example consider two images related only by a change of scale: the neighborhood of correspondent points will contain the same image portion only if it scales covariantly with the images. The only set of transformations where the structure of the neighborhood is preserved is the class of rotations and translation (the shape of a circle which is translated and rotated does not change). A discussion about automatic methods to retrieve correspondent neighborhoods in images related by transformation that do not preserve circular neighborhoods can be found in [6] and [7].

B. The Auto-Correlation Matrix

The smoothed version of the image $\mathcal{I}$ is given by:

$$L(\mathbf{q}, \sigma, \mathcal{I}) = (G_\sigma * \mathcal{I})(\mathbf{q})$$

where G_σ is a Gaussian function with standard deviation σ. The gradient (indicated using the symbol $\nabla_{\mathbf{x}}$) of L is used to compute the *auto-correlation matrix* associated to image $\mathcal{I}$:

$$\begin{aligned} \mu(\mathbf{q}, \sigma_I, \sigma_D, \mathcal{I}) &= \\ &\left(w_{\sigma_I} * \nabla_{\mathbf{x}} L(\cdot, \sigma_D, \mathcal{I}) \nabla_{\mathbf{x}}^T L(\cdot, \sigma_D, \mathcal{I})\right)(\mathbf{q}) = \\ &\int_{\mathbb{R}^2} w_{\sigma_I}(\mathbf{q} - \mathbf{x}) \nabla_{\mathbf{x}}^T L(\mathbf{x}, \sigma_D, \mathcal{I}) \nabla_{\mathbf{x}} L(\mathbf{x}, \sigma_D, \mathcal{I}) d\mathbf{x} \end{aligned} \quad (1)$$

where w_σ is a weighting function whose size is dependent on the parameter σ_I. We shall refer to σ_I as the integration scale (it is the parameter that defines size of the window about the point $\mathbf{q}'$), and to σ_D as the differentiation scale. The discretized version of equation (1) is:

$$\begin{aligned} \mu(\mathbf{q}, \sigma_I, \sigma_D, \mathcal{I}) &\approx \\ &\sum_{\mathbf{x}_i} w_{\sigma_I}(\mathbf{q} - \mathbf{x}_i) \nabla_{\mathbf{x}}^T L(\mathbf{x}_i, \sigma_D, \mathcal{I}) \nabla_{\mathbf{x}} L(\mathbf{x}_i, \sigma_D, \mathcal{I}) \end{aligned} \quad (2)$$

The vectors $\mathbf{x}_i$ are the points that form the discretized support set[1] of the weighting function w_{σ_I}. Expression (2) can be rewritten compactly in matrix form as:

$$\mu(\mathbf{q}, \sigma_I, \sigma_D, \mathcal{I}) \approx A^T W A \quad (3)$$

[1]Often times w_{σ_I} is chosen to be a Gaussian function, whose support is the entire plane. In this case the right hand side of equation (2) becomes an infinite summation. However, to deal with finite dimensional matrices, we can consider the support of w_{σ_I} to be the compact set $\{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x}^T \Sigma^{-1} \mathbf{x} \leq c\}$, where c is some suitable constant. Discretizing the previous set we obtain a finite summation.

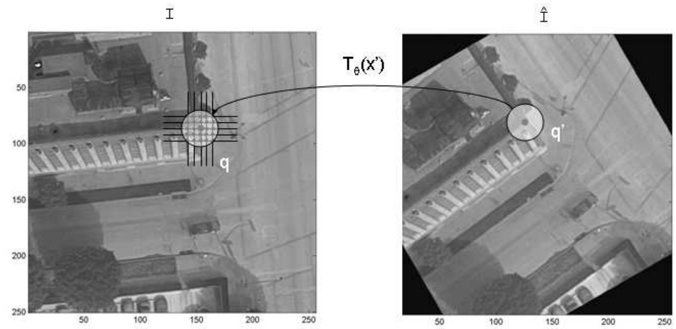

Fig. 1. The figure shows the mapping between the image pair $\mathcal{I}$ and $\hat{\mathcal{I}}$. The shaded circle represents the neighborhood of the point $\mathbf{q}$ and the black dots inside the shaded circle in the left image represent the points in the neighborhood of $\mathbf{q}$ that are used to construct the matrix A (see equation (3)).

where the matrix $A \in \mathbb{R}^{N \times 2}$ is formed by stacking the gradients $\nabla_{\mathbf{x}} L(\mathbf{x}_i, \sigma_D, \mathcal{I})$ one on top of the other and W is a diagonal weighting matrix such that $W_{ii} = w_{\sigma_I}(\mathbf{q} - \mathbf{x}_i)$. The matrix $A = A(\mathbf{q}, \sigma_D, \mathcal{I})$ depends both on the point $\mathbf{q}$ and on the differentiation scale σ_D. Similarly the matrix $W = W(\sigma_I)$ depends on the integration scale σ_I. However, to simplify the notation we will henceforth avoid to write explicitly such dependencies.

The auto-correlation matrix provides an important information about the image neighborhood for which the window function is non zero. Each eigenvalue μ measures the image gradient strength along the directions of the eigenvectors of μ. The larger the gradient strength the stronger the corner-like structure is.

C. Matrix Norms

In this section we will introduce the notation and the basic concepts about matrix norms that are used in this paper. For a more thorough discussion of these topics please refer to [8], [9]. Let's first introduce the vector *p-norm*:

$$\|\mathbf{x}\|_p \stackrel{\mathrm{def}}{=} \left(\sum_i x_i^p \right)^{\frac{1}{p}}$$

Vector p-norms lead to the definition of the *induced matrix p-norm*:

$$\|A\|_p \stackrel{\mathrm{def}}{=} \sup_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \quad (4)$$

It can be shown that $\|A\|_2 = \sigma_{max}(A)$, where $\sigma_{max}(A)$ is the maximum singular value of the matrix A. The *Schatten matrix p-norm* is defined as:

$$\|A\|_{S,p} \stackrel{\mathrm{def}}{=} \left(\sum_i \sigma_i(A)^p \right)^{\frac{1}{p}} \quad (5)$$

where $\sigma_i(A)$ is the i^{th} singular value of the matrix A.

III. Condition Theory for Point Matching

In this section we will introduce the fundamental facts about condition theory for point matching. A more thorough discussion of this theory can be found in [4]. The basic idea is to identify which points in an image can be used to estimate robustly the parameters that define a certain transformation. Suppose $\mathcal{I}$ and $\mathcal{I}'$ are two images related by the transformation $\mathbf{T}_{\boldsymbol{\theta}}$ so that $\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}}(\mathbf{q}')) = \mathcal{I}'(\mathbf{q}')$, and let's define the transformed image plus noise as:

$$\hat{\mathcal{I}}(\mathbf{q}') \stackrel{\text{def}}{=} \mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}}(\mathbf{q}')) + \eta(\mathbf{q}')$$

Let's also introduce the error cost function:

$$J_{\mathbf{T}_{\boldsymbol{\theta}}}(\mathbf{q}') \stackrel{\text{def}}{=} \frac{1}{2}\int_{\mathbb{R}^2} w_\sigma(\mathbf{q}'-\mathbf{x}')\left[\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}}(\mathbf{x}')) - \hat{\mathcal{I}}(\mathbf{x}')\right]^2 d\mathbf{x}' \quad (6)$$

In the noise free case (i.e. $\eta = 0$) the minimizer for (6) is given by $\boldsymbol{\theta}^*$ whereas in presence of noise such minimizer is moved to $\boldsymbol{\theta}^* + \Delta\boldsymbol{\theta}^*$. We would like to quantify the effect of the noise on the estimation of the transformation parameter $\boldsymbol{\theta}$. To achieve this goal we define a number that relates $\Delta\boldsymbol{\theta}$ to η in the limit for the noise tending to zero.

Definition 1: The *transformation condition number* (TCN) at point $\mathbf{q}'$ with respect to the transformation $\mathbf{T}_{\boldsymbol{\theta}}$ is defined as:

$$K_{\mathbf{T}_{\boldsymbol{\theta}}}(\mathbf{q}') \stackrel{\text{def}}{=} \lim_{\delta\to 0} \sup_{\|\eta\|\le\delta} \frac{\|\Delta\boldsymbol{\theta}^*\|}{\|\boldsymbol{\eta}\|} \quad (7)$$

where $\|\boldsymbol{\eta}\|$ takes into consideration the noise over the support of the window function w_σ.

The interpretation of expression (7) is the following: a large condition number means that small perturbations in the luminance of the image can greatly affect the value of the minimizer of (6). Therefore it is reasonable to seek a set of tie points for which the TCN has a small value. Unfortunately for an arbitrary transformation the TCN depends on the parameters of the transformation itself, so it is of little use if we want to assess the suitability of certain point to estimate $\boldsymbol{\theta}$ by using just one single image: in fact at this time $\boldsymbol{\theta}^*$ is unknown. This consideration leads us to a further definition.

Definition 2: The *class condition number* (CCN) at point $\mathbf{q}'$ with respect to the class of transformations $\mathcal{T}$ is defined as:

$$K_{\mathcal{T}}(\mathbf{q}') \stackrel{\text{def}}{=} \max_{\mathbf{T}_{\boldsymbol{\theta}}\in\mathcal{T}} K_{\mathbf{T}_{\boldsymbol{\theta}}}(\mathbf{q}') \quad (8)$$

In this case the condition number is independent from the transformation parameters. The following theorems provide a mean to calculate the condition numbers. For the proofs see the appendices I and II.

Theorem 1: The CCN at point $\mathbf{q}'$ for the class of translations $\mathcal{T}_T$ is approximated by:

$$K_{\mathcal{T}_T}(\mathbf{q}') = \left\|\left(A^TWA\right)^{-1}A^TW\right\| \quad (9)$$

where the matrices A and W are defined as in (3).

Theorem 2: For $W = I$, using the induced matrix 2-norm in (9), the CCN satisfy the following inequality:

$$K_{\mathcal{T}_T}(\mathbf{q}') \le K_{\mathcal{T}_{RT}}(\mathbf{q}')$$

Moreover, the scalar $\|\left(A^TA\right)^{-1}\|_2^2$ is a lower bound for $K_{\mathcal{T}_T}(\mathbf{q}')$ and $K_{\mathcal{T}_{RT}}(\mathbf{q}')$ for any weighting matrix W.

Theorem 1 provides a closed form expression for the CCN with respect to the class of translation transforms. Theorem 2 states that if a point is badly conditioned with respect to the class of transformations $\mathcal{T}_T$ it will also be bad conditioned with respect to the class of transformations $\mathcal{T}_{RT}$. This result agrees with the intuition: if we have to estimate more parameters using the same amount of data, then the noise will have a stronger influence on the estimate, producing a larger variation $\Delta\boldsymbol{\theta}^*$. The lower bound introduced in Theorem 2 provides a computationally efficient method to discard points that are bad conditioned: its calculation requires the computation of a matrix product and the solution of a second order equation to detect the minimum eigenvalue.

IV. Point Detectors

A point detector is an algorithm that takes one image as input and then outputs a set of *tie points* that can be identified with high repeatability in images that are related by a transformation $\mathbf{T}_{\boldsymbol{\theta}}$. Such an algorithm is completely defined as long as we specify the *rule* for detecting the tie points. In this section we will briefly describe the Harris-Stephens corner detector rule [1] (with a particular emphasis on the modification proposed by Noble [2]), the Shi-Tomasi point detector rule [3] and finally the point detector rule proposed by Kenney et al. in [4]. In all these works the detection rule is strongly connected to the spectral structure of the autocorrelation matrix μ. In general a point $\mathbf{q}$ is considered to be a tie-point as long as:

$$\mathcal{M}\left(\lambda\left(\mu(\mathbf{q}, \sigma_I, \sigma_D, \mathcal{I})\right)\right) \ge T \quad (10)$$

where $\mathcal{M}$ is a function of the eigenvalues of the autocorrelation matrix (which are indicated using the notation $\lambda(\cdot)$) and T is some suitable threshold.

As discussed at the end of section II, the strength of corner-like image structures is reflected in the magnitude of the eigenvalues of μ. Therefore it is natural to require $\mathcal{M}$ to be a non decreasing function of the eigenvalues of μ (in other words the stronger is a corner-like structure the stronger must be the response of the function 10). We will refer to this condition as the *nondecreasing constraint* on $\mathcal{M}$:

Constraint 1: Any function $\mathcal{M}$ used to measure the strength of a corner-like image structure based on the autocorrelation matrix A must be a non-decreasing function of the eigenvalues of A. In other words, for any $\lambda_1 \le \lambda_1'$, $\lambda_2 \le \lambda_2'$:

$$\mathcal{M}(\lambda_1, \lambda_2) \le \mathcal{M}(\lambda_1', \lambda_2') \quad (11)$$

A. Harris-Stephens and Noble Corner Detector

Harris and Stephens's corner detector draws its origins in the corner detector proposed by Moravec [10], where the authors consider to be good corners those points for which the difference between their neighborhood and shifted version of the same neighborhood produces an error surface with a well defined minimum. This idea encapsulates the notion of computational stability that is the core of the condition theory presented in [4] and summarized in section III. Harris and Stephens proposed the following function to measure the corner strength:

$$\mathcal{M}_H(\mathbf{q}, \sigma_I, \sigma_D, \mathcal{I}) = \det\left(A^T W A\right) - \alpha \left(\text{trace}\left(A^T W A\right)\right)^2 \quad (12)$$

The larger the value of α the less sensitive is the detector to corner like structures. The major drawback of this corner detector is the presence of a parameter α that needs to be manually tuned. To overcome this difficulty Noble [2] proposed a modified version of the Harris function that does not contain any constant:

$$\mathcal{M}_N(\mathbf{q}, \sigma_I, \sigma_D, \mathcal{I}) = \frac{\det\left(A^T W A\right)}{\text{trace}\left(A^T W A\right) + \varepsilon} \quad (13)$$

The small constant ε is used to avoid a singular denominator in case of a rank zero auto-correlation matrix.

B. Shi-Tomasi Point Detector

In [3] Shi and Tomasi proposed a criterion to decide which points are suitable for tracking (see also [11]). Their idea is to select points for which the system that provides an estimate of the displacement from one frame to the other is numerically well conditioned. Also in this case we find a connection with the core ideas of the condition theory summarized in section III. The Shi and Tomasi function is defined as:

$$\mathcal{M}_S(\mathbf{q}, \sigma_I, \sigma_D, \mathcal{I}) = \lambda_{min}(A^T W A) \quad (14)$$

where $\lambda_{min}(A^T W A)$ indicates the smallest eigenvalue of the auto-correlation matrix.

C. Kenney et al. Point Detector

As discussed in section III, Theorem (1) and Theorem (2) suggest the following rule to detect tie points: select those points that have a small condition number with respect to the class $\mathcal{T}_T$. The detection rule can be formalized by defining the Kenney's function:

$$\mathcal{M}_{K,*}(\mathbf{q}, \sigma_I, \sigma_D, \mathcal{I}) = \frac{1}{\left\| \left(A^T W A\right)^{-1} A^T W \right\|_*^2}$$

where $*$ indicates which matrix norm has been used.

V. Relation Between Point Detectors

A. Equivalence of Point Detectors

As pointed out in section IV, both the Harris-Stephens detector (and consequently the Nobel detector) and the Shi-Tomasi detector somehow encapsulate a notion of computational stability. This fact suggests that the functions (13) and (14) should reflect in some way this notion. To support this observation we will prove a theorem that establishes a relation between the point detectors described before, showing that the Noble function, the Shi-Tomasi function and the Kenney function are equivalent modulo the choice of a suitable matrix norm.

Theorem 3: For $\varepsilon = 0$ and $W = I$:

$$\mathcal{M}_N = \mathcal{M}_{K,1-Schatten} \quad (15)$$

$$\mathcal{M}_S = \mathcal{M}_{K,2} \quad (16)$$

Proof: Suppose the spectrum of the matrix $A^T W A = A^T A$ is given by $\lambda_2 \geq \lambda_1 \geq 0$. We will first show the equivalence relation between the Harris-Stephens-Nobel function and Kenney's one when the 1-Schatten norm is used. Consider the matrix $M = (A^T A)^{-1} A^T$. First we will prove that:

$$\sigma_i^2(M) = \sigma_i((A^T A)^{-1}) \quad (17)$$

This results follows directly from the SVD decomposition of the matrices M and $(A^T A)^{-1}$:

$$(A^T A)^{-1} = (V \Sigma^T U^T U \Sigma V^T)^{-1} = V \Sigma^{-2} V^T$$

$$M = (A^T A)^{-1} A^T = V \Sigma^{-2} V^T V \Sigma^T U^T = V \begin{bmatrix} \Sigma^{-1} & 0 \end{bmatrix} U^T$$

where Σ^{-2} is the diagonal matrix containing the 2 singular values of $(A^T A)^{-1}$. From the definition of the Schatten norm (5) it follows that:

$$\|(A^T A)^{-1}\|_{S,p} = \left[\sum_i \sigma_i((A^T A)^{-1})^p\right]^{\frac{1}{p}} = \left[\sum_i \sigma_i(M)^{2p}\right]^{\frac{1}{p}} = \left[\sum_i \sigma_i(M)^{2p}\right]^{\frac{1}{2p}\cdot 2} = \|M\|_{S,2p}^2$$

Given that for a symmetric matrix the singular values coincide with the eigenvalues we have that:

$$\mathcal{M}_{K,1-Schatten} = \frac{1}{\|(A^T A)^{-1}\|_{S,1}} = \frac{1}{\|M\|_{S,2}^2} = \text{(because of (17))} = \frac{1}{\frac{1}{\lambda_1} + \frac{1}{\lambda_2}} = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} = \text{(for } \varepsilon = 0) = \mathcal{M}_N$$

which proves the first part of the theorem, since the Nobel function (13) can be rewritten as:

$$\mathcal{M}_N = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2 + \varepsilon}$$

Now let's consider the equivalence between the Shi and Tomasi function and Kenney's one. Since the 2-norm of a matrix coincides with the largest singular value, using the identity (17) we can write:

$$\|M\|_2^2 = \|(A^TA)^{-1}\|_2$$

Therefore the following equality holds:

$$\mathcal{M}_{K,2} = \frac{1}{\|M\|_2^2} = \frac{1}{\|(A^TA)^{-1}\|_2} = \lambda_1 = \mathcal{M}_S \tag{18}$$

which proves also the second equivalence. ■

B. The Nondecreasing Constraint

The nondecreasing constraint (11) introduced at the end of Section IV says that the detection function should be nondecreasing in λ_1 and λ_2. The following lemma shows that $\mathcal{M}_N$ and $\mathcal{M}_S$ satisfy such constraint, whereas for the Harris-Stephens function $\mathcal{M}_H$ the constraint holds true only when $\alpha \leq \frac{1}{4}$ (this fact provides a justification of the common habit of setting $\alpha = 0.04$).

Lemma 1: The functions $\mathcal{M}_N$, $\mathcal{M}_S$ (and consequently $\mathcal{M}_{K,1-Schatten}$ and $\mathcal{M}_{K,2}$) all satisfy the nondecreasing constraint (11). The function $\mathcal{M}_H$ satisfies (11) provided that $\alpha \leq \frac{1}{4}$.

Proof: The constraint (11) is satisfied as long as the partial derivatives of the function $\mathcal{M}$ are non negative:

$$\frac{\partial \mathcal{M}}{\partial \lambda_1} \geq 0 \qquad\qquad \frac{\partial \mathcal{M}}{\partial \lambda_2} \geq 0$$

For the Nobel function we have that the partial derivatives are given by:

$$\frac{\partial \mathcal{M}_N}{\partial \lambda_i} = \frac{\lambda_j(\lambda_j + \varepsilon)}{(\lambda_i + \lambda_j + \varepsilon)^2}$$

which shows that they are always nonnegative. Now assume that $\lambda_2 \geq \lambda_1 \geq 0$. For the Shi-Tomasi function we have that:

$$\frac{\partial \mathcal{M}_N}{\partial \lambda_i} = \begin{cases} 1 & \text{for } \lambda_i = \lambda_1 \\ 0 & \text{for } \lambda_i = \lambda_2 \end{cases}$$

Therefore also in this case the partial derivatives are always non negative. From Theorem 3 it follows that Kenney's point detector satisfies the nondecreasing constraint when $W = I$ and either the 2-norm or the Schatten 1-norm are used. Finally let's consider the Harris-Stephens function for which:

$$\frac{\partial \mathcal{M}_H}{\partial \lambda_i} = \lambda_j - 2\alpha(\lambda_i + \lambda_j)$$

The partial derivatives are non negative if and only if:

$$\alpha \leq \frac{\lambda_1}{2(\lambda_1 + \lambda_2)}$$

Since $\frac{\lambda_1}{2(\lambda_1+\lambda_2)} \geq \frac{1}{4}$ for any $\lambda_2 \geq \lambda_1 \geq 0$, we conclude that $\mathcal{M}_H$ will satisfy (11) if and only if $\alpha \leq \frac{1}{4}$. ■

VI. Conclusion

In this paper we reformulated the point detection rules proposed by Harris, Stephens and Nobel, by Shi and Tomasi and by Kenney using a common framework based on condition theory introduced in [4] and summarized in section III. We restricted our attention to rotation and translation since more complex transformation would require an automatic procedure to detect neighborhoods that vary covariantly with the transformation [7]. The central result is stated in Theorem 3, where we showed that the point detectors previously listed are equivalent modulo a suitable choice of a matrix norm, (as long as the weighting function is constant over its support). This results tells that the Harris, Stephens and Nobel corner detector and the feature detector proposed by Shi and Tomasi both select points that are suitable to estimate robustly the parameters of a translation between an image pair. We also introduced a constraint that should be satisfied by all the functions $\mathcal{M}$ used to detect tie points based on the auto-correlation matrix. We showed that all the detectors studied in this paper satisfy the previous constraint automatically, with the relevant exception of the Harris Stephens corner detector, for which it is necessary to force α to be less than $\frac{1}{4}$.

Future research directions aim at considering a more flexible definition of condition numbers, in order to be able to cope with a larger variety of transformations possibly without having to solve explicitly the problem of the detection of a covariant neighborhood. We would also like to study in more detail which is the role of the weighting function w_σ, in order to present an extension of these results that holds also in the case where $W \neq I$.

Acknowledgments. This project was supported in part by the following grants from the Office of Naval Research: ONR #N00014-02-1-0121 (Zuliani and Manjunath) and ONR #N00014-02-1-0318 (Kenney).

Appendix I
Proof of Theorem 1.

Proof: Our goal is to express the parameter vector $\boldsymbol{\theta}^*$ that minimizes the cost (6) in terms of the noise η. Suppose $\eta \neq 0$: then the minimizer $\boldsymbol{\theta}^*$ will move to $\boldsymbol{\theta}^* + \Delta\boldsymbol{\theta}^*$. A necessary condition for optimality is that $\nabla_{\boldsymbol{\theta}} J_{\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}}(\mathbf{q}') = 0$, i.e. :

$$\int_{\mathbb{R}^2} w_\sigma(\mathbf{q}' - \mathbf{x}') \left[\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}(\mathbf{x}')) - \hat{\mathcal{I}}(\mathbf{x}')\right] \nabla_{\mathbf{x}}\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}(\mathbf{x}')) J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}(\mathbf{x}') d\mathbf{x}' = 0 \tag{19}$$

where $\nabla_{\mathbf{x}}\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}(\mathbf{x}'))$ is the image gradient computed at the point $\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}(\mathbf{x}')$ and $J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')$ is the Jacobian of the transformation $\mathbf{T}_{\boldsymbol{\theta}^*}$. The Taylor truncated expansion of the last three factors in (19) is given by:

$$\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}(\mathbf{x}')) - \hat{\mathcal{I}}(\mathbf{x}') \approx \nabla_{\mathbf{x}}\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')) J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')\Delta\boldsymbol{\theta}^* - \eta(\mathbf{x}')$$

$$\nabla_{\mathbf{x}}\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}(\mathbf{x}')) \approx \nabla_{\mathbf{x}}\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')) + \Delta\boldsymbol{\theta}^{*T} J_{\boldsymbol{\theta}}^T\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}') H_{\mathbf{x}}\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}'))$$

where $H_{\mathbf{x}}\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}'))$ is the image Hessian. As far as the Jacobian is concerned we have that in the pure translation case:

$$J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}(\mathbf{x}') = J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}') = I \tag{20}$$

whereas in the case of rotation and translation:

$$J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*+\Delta\boldsymbol{\theta}^*}(\mathbf{x}') \approx J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}') + M\Delta\boldsymbol{\theta}^* = \left[\begin{array}{cc} R(\phi+\frac{\pi}{2})\Delta\mathbf{q} & I \end{array}\right] + \left[\begin{array}{cc} R(\phi)\Delta\mathbf{q} & 0 \end{array}\right]\Delta\boldsymbol{\theta}^*$$

The matrix $R(\alpha)$ is the matrix in $SO(2)$ that produces a rotation of an angle α and $\Delta\mathbf{q} = \mathbf{x}' - \mathbf{q}'$. By dropping the second order terms (i.e. the terms containing the products $\Delta\boldsymbol{\theta}^{*T}\Delta\boldsymbol{\theta}^*$ and $\Delta\boldsymbol{\theta}^{*T}\eta(\mathbf{x}')$), we obtain the new equation:

$$\int_{\mathbb{R}^2} w_\sigma(\mathbf{q}'-\mathbf{x}')\left[\nabla_{\mathbf{x}}\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}'))J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')\Delta\boldsymbol{\theta}^* - \eta(\mathbf{x}')\right] \nabla_{\mathbf{x}}\mathcal{I}(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}'))J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}') \approx 0 \tag{21}$$

Let's now consider the discretized version of equation (21):

$$\sum_{\mathbf{x}'} w_\sigma(\mathbf{q}'-\mathbf{x}') \left[\nabla_{\mathbf{x}}L(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}'),\sigma_D,\mathcal{I})J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')\Delta\boldsymbol{\theta}^* - \eta(\mathbf{x}')\right] \nabla_{\mathbf{x}}L(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}'),\sigma_D,\mathcal{I})J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}') \approx 0 \tag{22}$$

where the continuous derivatives are replaced by convolutions with the derivatives of a gaussian kernel with standard deviation σ_D. By setting $\mathbf{v}^T(\mathbf{x}') = L(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}'),\sigma_D,\mathcal{I})J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')$, (22) can be written as:

$$\left[\sum_{\mathbf{x}'}\mathbf{v}(\mathbf{x}')w_\sigma(\mathbf{q}'-\mathbf{x}')\mathbf{v}^T(\mathbf{x}')\right]\Delta\boldsymbol{\theta}^* \approx \sum_{\mathbf{x}'} w_\sigma(\mathbf{q}'-\mathbf{x}')\mathbf{v}(\mathbf{x}')\eta(\mathbf{x}') \tag{23}$$

By defining W as in (3) and:

$$\boldsymbol{\eta} = \left[\begin{array}{ccc} \eta(\mathbf{x}'_1) & \dots & \eta(\mathbf{x}'_n) \end{array}\right]^T$$

(23) can be written as:

$$A^T_{\mathbf{T}_{\boldsymbol{\theta}^*}}(\mathbf{q}')WA_{\mathbf{T}_{\boldsymbol{\theta}^*}}(\mathbf{q}')\Delta\boldsymbol{\theta}^* \approx A^T_{\mathbf{T}_{\boldsymbol{\theta}^*}}(\mathbf{q}')W\boldsymbol{\eta} \tag{24}$$

The expression for the TCN is obtained by solving in a least square sense the normal equation (24). For a pure translation A does not depend on $\boldsymbol{\theta}^*$ (because $J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}') = I$) and therefore:

$$K_{\mathcal{T}_T}(\mathbf{q}') = K_{\mathbf{T}_{\boldsymbol{\theta}}}(\mathbf{q}') = \left\|\left(A^TWA\right)^{-1}A^TW\right\|$$

provides also an expression for the CCN for translation. The matrix A is the same as in (3). For the rotation and translation case the Jacobian $J_{\boldsymbol{\theta}}\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')$ unfortunately depends on $\boldsymbol{\theta}^*$ and therefore only an expression for the TCN can be provided:

$$K_{\mathbf{T}_{\boldsymbol{\theta}}}(\mathbf{q}') = \left\|\left(A^T_{\mathbf{T}_{\boldsymbol{\theta}^*}}(\mathbf{q}')WA_{\mathbf{T}_{\boldsymbol{\theta}^*}}(\mathbf{q}')\right)^{-1}A^T_{\mathbf{T}_{\boldsymbol{\theta}^*}}(\mathbf{q}')W\right\|$$

■

APPENDIX II
PROOF OF THEOREM 2.

We first show the following lemmas.

Lemma 2: Let A be defined as in (3) and let $\mathbf{w} \in \mathbb{R}^N$. Let's also define the matrix $A_J \stackrel{\text{def}}{=} \left[\begin{array}{cc} \mathbf{w} & A \end{array}\right]$. Then the following inequality holds:

$$\|\left(A_J^TA_J\right)^{-1}A_J^T\|_2 \geq \|\left(A^TA\right)^{-1}A^T\|_2 \tag{25}$$

Proof: As shown in the proof of Theorem 3, if (25) holds, then such inequality can be rewritten as:

$$\|\left(A_J^TA_J\right)^{-1}\|_2 \geq \|\left(A^TA\right)^{-1}\|_2$$

or equivalently as:

$$\lambda_{min}\left(A_J^TA_J\right) \leq \lambda_{min}\left(A^TA\right) \tag{26}$$

Therefore the lemma will be proved if we are able to prove (26). For our purposes it is more convenient to consider the transpose of the matrices $A_J^TA_J$ and A^TA; in particular we have that:

$$\lambda(A_JA_J^T) = \lambda(A_J^TA_J) \cup \{\underbrace{0,\cdots,0}_{N-3}\} = \{\lambda'_3,\lambda'_2,\lambda'_1,\underbrace{0,\cdots,0}_{N-3}\}$$

where $\lambda'_3 \geq \lambda'_2 \geq \lambda'_1 \geq 0$ and analogously:

$$\lambda(AA^T) = \lambda(A^TA) \cup \{\underbrace{0,\cdots,0}_{N-2}\} = \{\lambda_2,\lambda_1,\underbrace{0,\cdots,0}_{N-2}\}$$

where $\lambda_2 \geq \lambda_1 \geq 0$. If we rewrite the matrix $A_JA_J^T$ as $\mathbf{w}\mathbf{w}^T + AA^T$ we can exploit the interlacing property of the eigenvalues (see in particular Theorem 8.1.8 p. 397 in [9]) according to which $\lambda'_1 \in [0,\lambda_1]$. Therefore we can conclude that:

$$\lambda_{min}\left(A_J^TA_J\right) = \lambda'_1 \leq \lambda_1 = \lambda_{min}\left(A^TA\right)$$

which proves the lemma. ■

Lemma 3: For any matrices A and W with compatible dimensions the following inequality holds true:

$$\|\left(A^TWA\right)^{-1}A^TW\|_2 \geq \|\left(A^TA\right)^{-1}A^T\|_2 \tag{27}$$

Proof: As showed in more detail in the proof of Theorem 3, the SVD decomposition of the matrix A is:

$$A = U\left[\begin{array}{c} \Sigma \\ 0 \end{array}\right]V^T$$

and the SVD decomposition of $M = \left(A^TA\right)^{-1}A^T$ is:

$$M = V\left[\begin{array}{cc} \Sigma^{-1} & 0 \end{array}\right]U^T$$

Because the matrix 2-norm is invariant with respect to orthogonal rotations we have: $\|M\|_2 = \left\|\left[\begin{array}{cc} \Sigma^{-1} & 0 \end{array}\right]\right\|_2 = \|\Sigma^{-1}\|_2$. Introducing the matrix:

$$\left[\begin{array}{cc} m_{11} & m_{12} \\ m_{21} & m_{22} \end{array}\right] = U^TWU$$

compatibly dimensioned with U, it is possible to show that[2]:

$$\left(A^T W A\right)^{-1} A^T W = V \left[\begin{array}{cc} \Sigma^{-1} & \Sigma^{-1} m_{11}^{-1} m_{12} \end{array} \right] U^T$$

and consequently:

$$\| \left(A^T W A\right)^{-1} A^T W \|_2 = \left\| \left[\begin{array}{cc} \Sigma^{-1} & \Sigma^{-1} m_{11}^{-1} m_{12} \end{array} \right] \right\|_2$$

At this point the lemma is proved, since:

$$\left\| \left[\begin{array}{cc} \Sigma^{-1} & \Sigma^{-1} m_{11}^{-1} m_{12} \end{array} \right] \right\|_2 \geq \left\| \left[\begin{array}{cc} \Sigma^{-1} & 0 \end{array} \right] \right\|_2$$

■

Using these results we can now prove Theorem 2.

Proof: In the proof of Theorem 1 we showed that for rotation and translation the rows of the matrix $A_{\mathbf{T}_{\boldsymbol{\theta}^*}}(\mathbf{q}')$ can be written as $\mathbf{v}^T(\mathbf{x}') = L(\mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}'), \sigma_D, \mathcal{I}) J_{\boldsymbol{\theta}} \mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')$, where the Jacobian $J_{\boldsymbol{\theta}} \mathbf{T}_{\boldsymbol{\theta}^*}(\mathbf{x}')$ has the form $\left[\begin{array}{cc} \mathbf{u} & I \end{array} \right]$. Therefore we can write:

$$A_{\mathbf{T}_{\boldsymbol{\theta}^*}}(\mathbf{q}') = \left[\begin{array}{cc} \mathbf{w} & A \end{array} \right]$$

If we rename $A_{\mathbf{T}_{\boldsymbol{\theta}^*}}(\mathbf{q}')$ with A_J then we can apply Lemma 2, showing directly that the TCN for rotation and translation is always greater or equal than the CCN for translation (provided $W = I$). Since Lemma 2 holds for any vector $\mathbf{w}$, then the result can be extended to the CCN for rotation and translation, yielding the desired inequality:

$$K_{\mathcal{T}_T}(\mathbf{q}') \leq K_{\mathcal{T}_{RT}}(\mathbf{q}')$$

The proof of the last part of the theorem is a straightforward application of Lemma 3. ■

REFERENCES

[1] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. of the 4th ALVEY vision conference*, M. M. Matthews, Ed., University of Manchester, England, Septemeber 1988, pp. 147–151.

[2] A. Noble, "Descriptions of image surfaces," Ph.D. dissertation, Department of Engineering Science, Oxford University, 1989.

[3] J. Shi and C. Tomasi, "Good features to track," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, Seattle, Washington, June 1994, pp. 593–600.

[4] C. Kenney, B. Manjunath, M. Zuliani, G. Hewer, and A. Van Nevel, "A condition number for point matching with application to registration and post-registration error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1437–1454, November 2003.

[5] C. Schmid, R. Mohr, and C. Bauckhage, "Comparing and evaluating interest points," in *Proc. of IEEE 6th International Conference on Computer Vision*, Bombay, India, January 1998, pp. 230–235. [Online]. Available: http://www.inrialpes.fr/movi/publi/Publications/1998/SMB98

[6] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Dordrecht, Netherlands: Kluwer Academic, 1994.

[7] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *European Conference on Computer Vision*. Copenhagen, Denmark: Springer, 2002, pp. 128–142. [Online]. Available: http://www.inrialpes.fr/movi/publi/Publications/2002/MS02

[8] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The John Hopkins University Press, 1996.

[9] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2001.

[10] H. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover," Carnegie-Mellon University, Robotics Institute, Tech. Rep. CMU-RI-TR-3, September 1980.

[11] W. Förstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and center of circular features," in *Proc. of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland*, June 2-4 1987, pp. 281–305.

[2]We assume m_{11} to be invertible: if it is not a simple limiting argument gives the same result.