Point Matching under Large Image Deformations and Illumination Changes

Bogdan Georgescu, *Student Member*, *IEEE*, and Peter Meer, *Senior Member*, *IEEE*

Abstract—To solve the general point correspondence problem in which the underlying transformation between image patches is represented by a homography, a solution based on extensive use of first order differential techniques is proposed. We integrate in a single robust M-estimation framework the traditional optical flow method and matching of local color distributions. These distributions are computed with spatially oriented kernels in the 5D joint spatial/color space. The estimation process is initiated at the third level of a Gaussian pyramid, uses only local information, and the illumination changes between the two images are also taken into account. Subpixel matching accuracy is achieved under large projective distortions significantly exceeding the performance of any of the two components alone. As an application, the correspondence algorithm is employed in oriented tracking of objects.

Index Terms—Correspondence problem, optical flow, color distribution matching, motion tracking, wide-baseline stereo.

1 Introduction

Finding corresponding features between different images of a scene is the first module in almost any 3D computer vision task, with the performance of the subsequent modules being conditioned by the accuracy of the matched features. In this paper, we focus on point correspondences, and also apply our technique to object tracking.

The correspondences are established by exploiting local photometric and/or geometric characteristics in the two images. To account for the local image deformations due to the change in the viewpoint, these deformations are approximated by a rigid transformation between corresponding image patches. The difficulty of matching increases with the complexity of the assumed local transformation. Three main processing steps can be distinguished in any correspondence algorithm: *feature detection*, in which salient interest point candidates are delineated; *feature matching*, in which the candidates are paired; and *match validation*, in which the most probable pairs are established.

A reliable *feature detection* method must remain stable under changes in the viewpoint and/or illumination. The Harris corner detector [24] was shown to have the most consistent performance over a large range of operating conditions [51], and is widely used for defining salient point features. Recently, it was generalized to tolerate affine deformations of an image neighborhood [39]. After the detection of the interest points, the correspondence between them can be established using one of two strategies.

The most popular methods for *feature matching* are correlation-based. Using the differential approach of optical flow, the correlation score is expressed as a function in the parameters of the local transformation between a reference patch in the first image and the current patch in the second image. An iterative maximization procedure then estimates these parameters, e.g., [34], [52], [6], [22], [2]. The two image patches can now be *registered* and, thus, implicitly their centers are put in correspondence.

The optical flow-based registration, however, has an important limitation. The employed optimization procedure requires that the two image patches are already similar at the pixel level. For example, large translations or rotations which are characteristic for wide-baseline stereo are not tolerated. Numerous techniques were proposed to alleviate this deficiency. A hierarchical framework, either isotropic [4] or anisotropic [65], is often used. The procedure can be also adapted for registration of images acquired with different sensors [27]. To assure an implicit smoothness of the flow field, directional regularization is proposed in [2], basis functions are used in [54], and mesh induced planar patches are matched in [21]. The estimation procedure can be enhanced in a robust framework for multiple motion [5], [46].

The optical flow-based matching can also be supported by using invariants. Affine invariant descriptors are built from local information [55], [39], [33] or are based on regions [35], [61]. In [39] iterative whitening of the local covariance matrix is used, the optimal size of the neighborhood being established through a search in the discrete scale space. The initial matches are based on the distance between descriptors derived from differentiation filters up to fourth order, and these matches are then verified using the correlation score. The approach was used in [48] for 3D object modeling and recognition. Local scale-invariant features derived from gradient orientation and magnitude are used in [33] also for object recognition. In [61], affine invariant regions are delineated starting from the local maxima of the image intensity and are bounded by extrema in the intensity variation. Correlation and invariant moment descriptors are used for matching. Intensity profiles between two points are used in [55] and, in [56], the method is extended to exploit the cyclical ordering of matched features in a wide-baseline stereo application. The effect of geometric transformation on template matching is explicitly taken into account in [3] by introducing the concept of geometric blur. However, to determine the adequate scale of the invariant descriptors in all these methods, often an explicit search in the parameter space

- *B. Georgescu is with the Real-Time Vision and Modeling Department, Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540. E-mail: georgesc@caip.rutgers.edu.*
- *P. Meer is with the Electrical and Computer Engineering Department, Rutgers University, 94 Brett Road, Piscataway, NJ 08854-8058. E-mail: meer@caip.rutgers.edu.*

Manuscript received 27 May 2003; revised 23 Sept. 2003; accepted 29 Sept. 2003. Recommended for acceptance by A. Rangarajan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0111-0603.

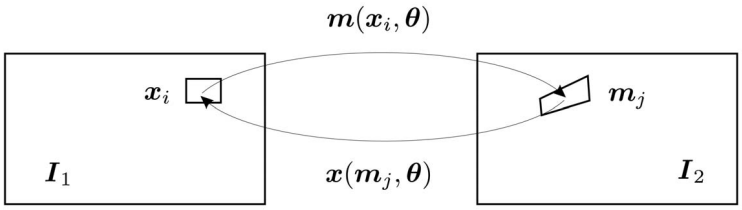

Fig. 1. Corresponding image patches.

is required. In a different approach toward matching, the point correspondences are established in a global process based on tensor voting using the four-dimensional space of the induced optical flow [45]. In [20], more complex features such as statistical distribution of geometric properties of image contours are used to estimate the transformation parameters.

Illumination changes between images further complicate the matching process and several techniques have been proposed to either derive illumination invariant features or to model the illumination changes and solve for the model parameters. In the context of optical flow, the dependence of performance on illumination changes is investigated in [57] where, also, a simple intensity mean substraction step is proposed to increase the performance of a tracking algorithm. In [23], basis images are used to determine the alignment difference due to illumination, while in [64], a method based on the localized consistency principle is described. The photometric changes are modeled in [40] and invariant features are derived based on generalized color moment invariants. In the context of image retrieval, color invariants are proposed in [19] and, in [12], a color channel normalization step is employed to reduce the influence of illumination changes.

The last module of a correspondence algorithm is *match validation*, which sometimes is combined with the matching step [47], [50]. The epipolar geometry induced by an image pair, or the trifocal tensor associated with a group of three images, provide global constraints which should be satisfied by all the correct matches. Because of the errors in the matching step, the geometric parameters have to be estimated robustly and methods such as RANSAC [16], MLESAC [59] or IMPSAC [58] are most often employed. Once a set of correct correspondences is found, the recovered geometry can be used to extend the number of matches under a constrained motion model. In [47], homographies are used to extend the matches, while in [50], matching is extended to multiple views by indexing invariant image patches and using geometric constraints between two and three views. It is important to emphasize that all the match validation procedures employing a robust estimator require that the absolute majority of the matches is *already* correct.

The point correspondence algorithm proposed in this paper avoids many of the limitations of the methods mentioned above. The emphasis is put on the five-dimensional nature of the color image registration problem, which is then approached in two complimentary ways. Beside the traditional optical flow technique (Section 2.2), we also introduce a method for matching the color distributions derived from the two image patches through spatially oriented kernels. Both methods use the same parametrization for the underlying transformation, and they are combined into a single estimation process which yields a better matching performance than each of its components (Section 2.4). By combining the two techniques, we succeeded to overcome the limitations of both components. Color distribution-based matching is robust but has relative poor localization accuracy. Optical flow-based registration, on the other hand, is very accurate but needs close prior alignment. In our method, the processing moves gradually from the former to the latter. The performance of the algorithm to find point correspondences under severe geometric distortions and illumination changes is illustrated with several examples in Section 3. In Section 4, the matching algorithm is integrated into a high accuracy oriented tracker. Issues related to the employed estimation technique and the uncertainty of the obtained matches are discussed in Section 5, while in Section 6, the proposed correspondence algorithm is put in the context of computer vision literature.

2 Point Matching in 5D

The information in a color image can be represented in the five-dimensional space of the two spatial coordinates and the three components of the employed color representation (RGB, Luv, etc.). In this section, after discussing the mapping between two images of a scene, two methods for establishing point correspondences are described, each approaching the 5D information differently. In both methods, the same parametrization is used for the transformation between the image patches and, therefore, they can be combined into a single process for estimating the parameters of the transformation.

2.1 Local Registration

Given two color images I_1 and I_2 of a visual scene, an arbitrary point dependent transformation exists between their pixels $\boldsymbol{x}_i$ and $\boldsymbol{m}_j$, respectively (Fig. 1). To locally approximate the relation between two image patches $I_1(\boldsymbol{x}_i)$ and $I_2(\boldsymbol{m}_j)$, we use the most general linear transformation between the homogeneous coordinates, the planar homography. The mapping between a neighborhood centered on $\boldsymbol{x}_0$ in the first image and a neighborhood centered on $\boldsymbol{m}_0$ in the second image is defined as

$$\begin{bmatrix} \boldsymbol{m} - \boldsymbol{m}_0 \\ 1 \end{bmatrix} \propto \begin{bmatrix} A & \mathbf{0} \\ \boldsymbol{v}^\top & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{x} - \boldsymbol{x}_0 \\ 1 \end{bmatrix}, \tag{1}$$

where $\propto$ denotes projective equivalence. The projective deformation is modeled by $\boldsymbol{v} = [v_0 \; v_1]^\top$, while A is the 2×2 matrix of an affine transformation. The matrix A can be further decomposed using two rotation matrices R_0, R_1 and an anisotropic scaling matrix S ([23], p. 19)

$$\mathrm{A} = \mathrm{R}_0 \mathrm{R}_1^\top \mathrm{S} \mathrm{R}_1. \tag{2}$$

The 2D rotation matrices are parametrized by the angles α_0, α_1, respectively, and the scaling matrix is $\mathrm{S} = \mathrm{diag}[s_x, s_y]$. The main advantage of the decomposition (2) is that it allows the inverse mapping from the neighborhood centered on $\boldsymbol{m}_0$ in the second image to the neighborhood centered on $\boldsymbol{x}_0$ in the first image to be expressed using the same parameters

$$\begin{bmatrix} \boldsymbol{x}-\boldsymbol{x}_0 \\ 1 \end{bmatrix} \propto \begin{bmatrix} \mathrm{A}^{-1} & \mathbf{0} \\ -\boldsymbol{v}^\top \mathrm{A}^{-1} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{m}-\boldsymbol{m}_0 \\ 1 \end{bmatrix}, \tag{3}$$

where $\mathrm{A}^{-1} = \mathrm{R}_1^\top \mathrm{S}^{-1} \mathrm{R}_1 \mathrm{R}_0^\top$. Thus, both transformations are parametrized by an eight-dimensional vector $\boldsymbol{\theta}$: the spatial coordinates m_{0x}, m_{0y}, and α_0, α_1, s_x, s_y, v_0, v_1. Note that $\boldsymbol{x}_0$ is known, being the center of the reference neighborhood in the first image, while $\boldsymbol{m}_0$, its corresponding location in the second image, is to be determined.

In the sequel, these transformations are denoted as

$$\begin{array}{lll} \boldsymbol{I}_1 \rightarrow \boldsymbol{I}_2 & \boldsymbol{m}_i = \boldsymbol{m}(\boldsymbol{x}_i, \boldsymbol{\theta}) & \boldsymbol{x}_i \in \mathcal{N}_{x_0},\ i = 1, \ldots, n_x \\ \boldsymbol{I}_2 \rightarrow \boldsymbol{I}_1 & \boldsymbol{x}_j = \boldsymbol{x}(\boldsymbol{m}_j, \boldsymbol{\theta}) & \boldsymbol{m}_j \in \mathcal{N}_{m_0},\ j = 1, \ldots, n_m, \end{array} \tag{4}$$

where $\mathcal{N}_{\boldsymbol{u}}$ stands for the neighborhood centered on $\boldsymbol{u}$. Note that the index of a mapped pixel (in general, not on the image lattice) is that of the grid point in the other image.

The goal is to find an estimate of $\boldsymbol{\theta}$ that best approximates the transformation between the two image patches. We will make extensive use of the differential approach in which starting from an initial value $\hat{\boldsymbol{\theta}}^0$, the estimate $\hat{\boldsymbol{\theta}}$ is iteratively refined. For alignment using the differential method, is was shown in [1] that updating the parameter estimate (additive approach), is equivalent to updating the image (compositional approach). In the absence of additional information, the initial parameter value $\hat{\boldsymbol{\theta}}^0$ represents only a translation, that is, we are seeking an initial location $\boldsymbol{m}_0$ to assure the overlap between the corresponding image patches.

2.2 Optical Flow-Based Registration

In optical flow methods, the error is measured on the image grid between a reference image patch and the transformed patch from the second image. The mapping error between the images $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ is the color difference between the corresponding pixels

$$\boldsymbol{e}_{of}(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}}) = \boldsymbol{I}_2[\boldsymbol{m}(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}})] - \boldsymbol{I}_1(\boldsymbol{x}_i). \tag{5}$$

To obtain an update rule for the current parameter estimate $\hat{\boldsymbol{\theta}}$, i.e., to compute $\hat{\boldsymbol{\theta}}' = \hat{\boldsymbol{\theta}} + \delta\boldsymbol{\theta}$, the first order Taylor expansion of the error around $\hat{\boldsymbol{\theta}}$ is used

$$\boldsymbol{e}_{of}(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}}') \approx \boldsymbol{I}_2(\boldsymbol{m}_i) + \mathrm{J}_{I_2|\hat{\boldsymbol{\theta}}}^\top \delta\boldsymbol{\theta} - \boldsymbol{I}_1(\boldsymbol{x}_i), \tag{6}$$

where $\mathrm{J}_{I_2|\hat{\boldsymbol{\theta}}}^\top$ is the 8x3 Jacobian matrix computed in $\hat{\boldsymbol{\theta}}$ having as elements the derivatives of each color component with respect to each transformation parameter. See Appendix C for the definition of the Jacobian matrix.

To estimate $\hat{\boldsymbol{\theta}}$, the error in (6) is minimized over all pixels $\boldsymbol{x}_i$, $i = 1 \ldots n_x$ in the neighborhood $\mathcal{N}_{x_0}$ centered on $\boldsymbol{x}_0$. For a robust behavior of the estimation process, a biweight M-estimator is employed *and* the contribution of a pixel is also weighted with a spatial kernel according to its distance from the center $\boldsymbol{x}_0$ of the neighborhood. The procedure becomes a generalized M-estimation method. Appendix B provides an overview of M-estimators.

The optimization criterion thus can be expressed from (5) and (6) as

$$\mathcal{J}_{of} = \sum_{i=1}^{n_x} K_e\Big(\frac{\boldsymbol{x}_i - \boldsymbol{x}_0}{h}\Big) \rho\left(\left\| \frac{\mathrm{J}_{I_2|\hat{\boldsymbol{\theta}}}^\top \delta\boldsymbol{\theta} + \boldsymbol{e}_{of}(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}})}{\sigma} \right\| \right), \tag{7}$$

where h represents the spatial bandwidth and σ the scale of the color error, K_e is the radially symmetric Epanechnicov kernel (A.4), and $\rho(u)$ is the biweight loss function (B.4). For the multivariate case, the loss function is computed through the norm of the argument (Appendix B). In our experiments, $h = 35$ pixels and σ is 20 percent of the maximum range of color difference, which is the length of the main diagonal in the RGB color cube. The neighborhood should contain sufficient local information and the choice of the scale σ ensures that large errors in the first order approximation do not influence the estimate.

The optimization problem (7) is solved by iterative weighted least squares (B.8) and it can be shown that the parameter update equation is

$$\widehat{\delta\boldsymbol{\theta}} = -\left[\sum_{i=1}^{n_x} w_i \mathrm{J}_{I_2|\hat{\boldsymbol{\theta}}} \mathrm{J}_{I_2|\hat{\boldsymbol{\theta}}}^\top \right]^+ \left[\sum_{i=1}^{n_x} w_i \mathrm{J}_{I_2|\hat{\boldsymbol{\theta}}} \boldsymbol{e}_{of}(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}}) \right]. \tag{8}$$

The least squares estimate provides a stable solution even though it assumes a simplified noise model. See Section 5 for an in depth discussion of the related issues. The expression of the weights w_i is the product between a fixed weight from the kernel K_e and a variable weight from the loss function $\rho(u)$ (B.9)

$$w_i = K_e\Big(\frac{\boldsymbol{x}_i - \boldsymbol{x}_0}{h}\Big) \cdot \left(\frac{1 - \boldsymbol{e}^\top \boldsymbol{e}}{\sigma^2}\right)^2 \quad \|\boldsymbol{e}\| \leq \sigma. \tag{9}$$

The Jacobian in the image $\boldsymbol{I}_2$ with respect to the parameter $\boldsymbol{\theta}$ is computed by the chain rule (C.2)

$$\mathrm{J}_{I_2|\hat{\boldsymbol{\theta}}} = \mathrm{J}_{m_i|\hat{\boldsymbol{\theta}}} \mathrm{J}_{I_2|m_i}, \tag{10}$$

where $\mathrm{J}_{I_2|m_i}$ is the 2×3 matrix having as columns the gradient of each color component, and $\mathrm{J}_{m_i|\hat{\boldsymbol{\theta}}}$ is the Jacobian of the transformed coordinates $\boldsymbol{m}$ with respect to the parameter $\boldsymbol{\theta}$, computed analytically from (1). Appendix C provides an outline of the Jacobian computation.

The three color plane gradients in the second image are computed with 11×11 smoothed differentiation filters. The expressions for the separable filter sequences for a neighborhood of size $2n+1$ are

$$\begin{array}{ll} h_s(i) = \frac{1}{2^{2n}} \begin{pmatrix} 2n \\ n+i \end{pmatrix} & i = -n, \ldots, n \quad \text{smoothing} \\ h_d(i) = \frac{2i}{n} h_s(i) & i = -n, \ldots, n \quad \text{differentiation.} \end{array} \tag{11}$$

These filters are built using orthogonal polynomial bases and are optimal in the least squares sense [38].

The optical flow-based registration is well-known and is widely used in computer vision, e.g., [34], [52]. Its main deficiency is also well-known. To obtain an accurate alignment between the two neighborhoods (to match their centers), the neighborhoods must already have a significant alignment *prior* to the parameter estimation. This limitation is illustrated in Fig. 2, where a large deformation exist between the reference image patch (Fig. 2b) and the initial neighborhood in the second image (Fig. 2d). Using only optical flow for registration fails to recover the transformation between these two image patches.

2.3 Matching Color Distributions

The registration method discussed in Section 2.2 does not exploit all the information available in the color space. To use this information, we generalize a technique proposed for tracking in [11]. The color information of the neighborhood in the first image will be described by the *discrete* color density distribution $\boldsymbol{p}$

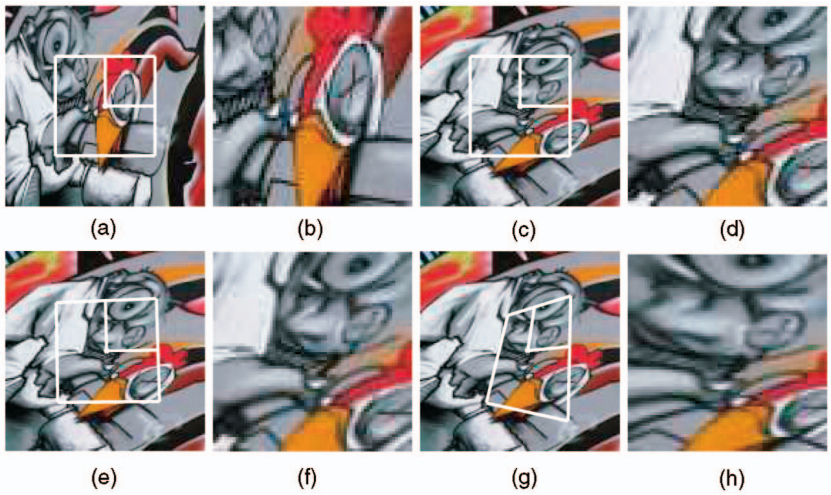

Fig. 2. *Alignment through optical flow.* The reference neighborhood delineated in the first image (a) is shown in (b). The neighborhood delineated at the same location in the second image (c) is shown in (d). The transformed neighborhood in the second image after two iterations, (e) and (f), and at convergence (g) and (h). The alignment is not successful.

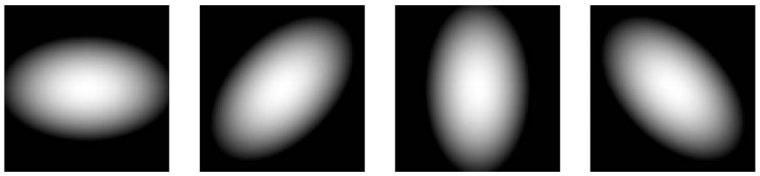

Fig. 3. Epanechnikov kernels oriented at 0^o, 45^o, 90^o, and 135^o.

$$\boldsymbol{p} = \{p_u\}_{u\in\mathcal{B}} \qquad \sum_{u\in\mathcal{B}} p_u = 1, \tag{12}$$

where $\mathcal{B}$ is a sampling of the employed color space. We used uniform sampling with $b = 16$ quantization steps per color coordinate and, thus, $\mathcal{B}$ has $16^3 = 4,096$ samples.

The distribution $\boldsymbol{p}$ is derived from the data using kernel density estimation, a technique reviewed in Appendix A. The expression of the density for each bin is

$$p_u = C_p \sum_{i=1}^{n_x} K_e\Big(\frac{\boldsymbol{c}_i - \boldsymbol{c}_u}{b}\Big), \tag{13}$$

where $\boldsymbol{c}_i$ represents the color of the pixel $\boldsymbol{x}_i$, $i = 1 \ldots n_x$, and $\boldsymbol{c}_u$ the color associated with the uth bin in $\mathcal{B}$. Taking the kernel bandwidth equal to the quantization step b means that the pixel i contributes only to the bins within unit distance. Again, the Epanechnikov kernel (A.4) is employed and the normalization constant C_p is such that (12) is satisfied.

As defined above, the distribution $\boldsymbol{p}$ is not sensitive to rotations (up to quantization effects) which is not desirable in applications involving alignments. To make the color distribution dependent on the deformation of the image patch, the contribution of each pixel's color is weighted with a value dependent on the pixel location. Therefore, an additional *elliptically symmetric* kernel $K_{e\beta}$ (A.5) is introduced in the spatial domain and (13) becomes

$$p_{u\beta} = C_p \sum_{i=1}^{n_x} K_{e\beta}\Big(\frac{\boldsymbol{x}_i - \boldsymbol{x}_0}{h}\Big) K_e\Big(\frac{\boldsymbol{c}_i - \boldsymbol{c}_u}{b}\Big), \tag{14}$$

where h is the size of the spatial neighborhood. Four kernels, oriented at $\beta = 0^o$, 45^o, 90^o, and 135^o suffice to cover the entire neighborhood (Fig. 3).

The local spatial structure in the second image is connected to that in the first image through the transformation parametrized by $\boldsymbol{\theta}$ (4). Instead of deforming the neighborhood in the second image, we can deform the support of the kernel $K_{e\beta}$ according to this transformation. Thus, the full color information from the second image is used directly and the image warping errors due to interpolation are avoided. The spatial relation between the support of kernel $K_{e\beta}$ and the neighborhood in the second image can now be obtained by mapping $K_{e\beta}$ through the *inverse* transformation (4).

The color distribution of the pixels $\boldsymbol{m}_j$, $j = 1 \ldots n_m$, in the neighborhood $\mathcal{N}_{m_0}$ centered on $\boldsymbol{m}_0$ in the second image are defined as

$$q_{u\beta}(\boldsymbol{\theta}) = C_q \sum_{j=1}^{n_m} K_{e\beta}\Big(\frac{\boldsymbol{x}(\boldsymbol{m}_j, \boldsymbol{\theta}) - \boldsymbol{x}_0}{h}\Big) K_e\Big(\frac{\boldsymbol{c}_j - \boldsymbol{c}_u}{b}\Big), \tag{15}$$

where C_q is the normalization constant such that $\sum_{u\in\mathcal{B}} q_{u\beta} = 1$.

The error between two components of the color distribution, computed with the kernel oriented β, will be measured by

$$e_{cd}(u, \beta, \boldsymbol{\theta}) = \sqrt{q_{u\beta}(\boldsymbol{\theta})} - \sqrt{p_{u\beta}} \qquad u \in \mathcal{B} \tag{16}$$

since, then, a least squares minimization is equivalent to maximizing the correlation coefficient between the two distributions

$$\operatorname*{argmin}_{\boldsymbol{\theta}} \sum_{u\in\mathcal{B}} e^2_{cd}(u, \beta, \boldsymbol{\theta}) \quad = \quad \operatorname*{argmax}_{\boldsymbol{\theta}} \sum_{u\in\mathcal{B}} \sqrt{q_{u\beta}(\boldsymbol{\theta}) p_{u\beta}}\ . \tag{17}$$

Recall that $\sum_{u\in\mathcal{B}} p_{u\beta} = \sum_{u\in\mathcal{B}} q_{u\beta}(\boldsymbol{\theta}) = 1$.

Similarly to optical flow-based registration, we use the first order approximation of the error (16) around the current estimate of the parameters $\hat{\boldsymbol{\theta}}$

$$e_{cd}(u, \beta, \hat{\boldsymbol{\theta}}') \approx \sqrt{q_{u\beta}(\hat{\boldsymbol{\theta}})} + \boldsymbol{g}^{\top}_{\sqrt{q_{u\beta}}|\hat{\boldsymbol{\theta}}}\, \delta\boldsymbol{\theta} - \sqrt{p_{u\beta}}, \tag{18}$$

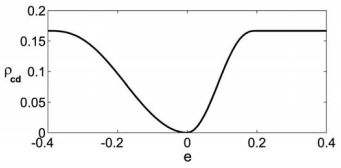

Fig. 4. The asymmetric biweight loss function $\rho_{cd}(u)$.

where $\boldsymbol{g}^{\top}_{\sqrt{q_{u\beta}}|\hat{\theta}}$ is the gradient of $\sqrt{q_{u\beta}}$ with respect to the parameter $\boldsymbol{\theta}$, computed in the current estimate $\hat{\boldsymbol{\theta}}$.

To obtain a robust solution for the minimization problem, the employed loss function should take into account the asymmetric nature of the task. Indeed, we should penalize more the case in which the color $p_{u\beta}$ from the reference distribution is not well represented in $q_{u\beta}(\boldsymbol{\theta})$, i.e., $p_{u\beta} >> q_{u\beta}(\boldsymbol{\theta})$, than the opposite case when $p_{u\beta} << q_{u\beta}(\boldsymbol{\theta})$.

The asymmetry is captured by switching the value of the scale σ depending on the sign of the error e_{cd} (16). The optimization criterion thus is defined as

$$\mathcal{J}_{cd} = \sum_{\beta} \sum_{u \in \mathcal{B}} \rho_{cd} \left(\frac{\boldsymbol{g}^{\top}_{\sqrt{q_{u\beta}}|\hat{\theta}} \delta\boldsymbol{\theta} + e_{cd}(u, \beta, \hat{\boldsymbol{\theta}})}{\sigma_{\pm}} \right), \tag{19}$$

where the loss function ρ_{cd} is based on the biweight (B.4)

$$\rho_{cd}\left(\frac{e}{\sigma_{\pm}}\right) = \begin{cases} \rho(e/\sigma_{+}) & e \geq 0 \\ \rho(e/\sigma_{-}) & e < 0 \end{cases} \tag{20}$$

and $\sigma_{+} = 0.2$, $\sigma_{-} = 0.4$ of the maximum range of color error (Fig. 4).

The optimization criterion (19) is again solved through iterative weighted least squares (B.8), and the update for the current parameter estimate $\hat{\boldsymbol{\theta}}$ is

$$\widehat{\delta\boldsymbol{\theta}} = -\left[\sum_{\beta} \sum_{u \in \mathcal{B}} w_u \boldsymbol{g}_{\sqrt{q_{u\beta}}|\hat{\theta}} \boldsymbol{g}^{\top}_{\sqrt{q_{u\beta}}|\hat{\theta}} \right]^{+} \left[\sum_{\beta} \sum_{u \in \mathcal{B}} w_u \boldsymbol{g}_{\sqrt{q_{u\beta}}|\hat{\theta}} e_{cd}(u, \beta, \hat{\boldsymbol{\theta}}) \right]. \tag{21}$$

The weights w_u are derived from $\rho_{cd}(u)$ and are similar to the second term in (9)

$$w_u = \left(\frac{1 - \boldsymbol{e}^{\top}\boldsymbol{e}}{\sigma_{\pm}^2} \right)^2 \qquad \|\boldsymbol{e}\| \leq \sigma_{\pm} \,. \tag{22}$$

The gradient is computed applying the chain rule (C.2) to (15)

$$\boldsymbol{g}_{\sqrt{q_{u\beta}}|\hat{\theta}} = \frac{1}{2} \frac{1}{\sqrt{q_{u\beta}}} C_q \sum_{j=1}^{n_m} \mathrm{J}_{x_j|\hat{\theta}} \boldsymbol{g}_{K_{e\beta}|x_j} K_e\left(\frac{\boldsymbol{c}_j - \boldsymbol{c}_u}{b}\right), \tag{23}$$

where the gradient of the kernel $K_{u\beta}$, computed analytically, is

$$\boldsymbol{g}_{K_{e\beta}|x_j} = -2h^{-1}\mathrm{B}_{\beta}^{-1}(\boldsymbol{x}_j - \boldsymbol{x}_0) \tag{24}$$

and $\mathrm{J}_{x_j|\hat{\theta}}$ is the Jacobian matrix of the inverse transform computed analytically from (3) (see Appendix C).

To illustrate the performance of the point matching technique introduced in this section, the example used for the optical flow (Fig. 2) is revisited. Now, after two iterations, the current parameter estimate transforms the spatial kernel in the second image as shown in Fig. 5e. When the region covered by the kernel in the second image is mapped into the first image coordinate system (Fig. 5f), it already shows a reasonable alignment. Recall that translations are estimated through $\boldsymbol{m}_0$. The estimation process converges after seven iterations, and the result (Fig. 5h) exhibits a good alignment (Fig. 5b). Compare with Fig. 2h where the optical flow failed to register the two image patches. However, as it will be seen in the next section, the location estimate is still off by about two pixels due to the fact that the color distributions are not very sensitive to the exact grid location.

2.4 Joint Estimation

The two point matching methods described in the previous sections work in the same five-dimensional space. Their strength and weaknesses are complimentary. Optical flow-based registration emphasizes the role of the sampling grid transformation and, therefore, can have superior localization accuracy. To be effective, however, the estimation process must start from a significant overlap between

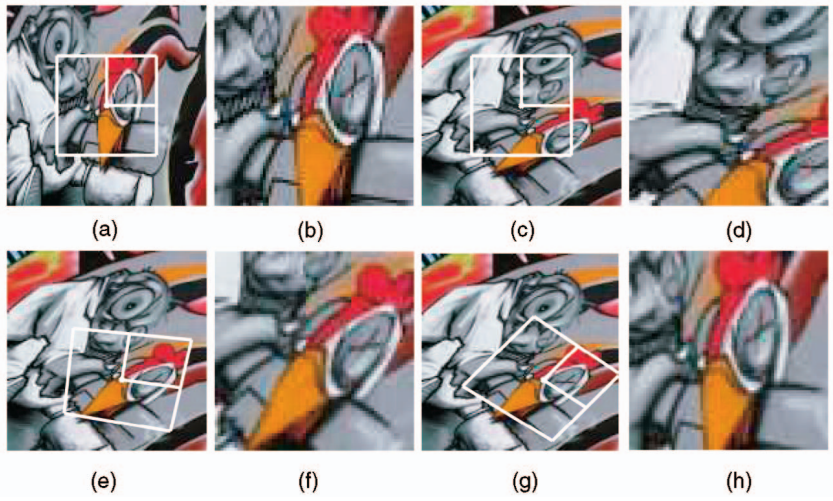

Fig. 5. *Alignment through matching color distributions*. The reference neighborhood delineated in the first image (a) is shown in (b). The neighborhood delineated at the same location in the second image (c) is shown in (d). The transformed neighborhood in the second image after two iterations, (e) and (f), and at convergence (g) and (h). The alignment appears to be satisfactory.

(a) (b) (c)

Fig. 6. *Performance comparison between different alignment strategies.* Registration error, shown as pixelwise image difference, based on (a) optical flow only, (b) color distribution only, and (c) joint estimation.

corresponding pixels in the two neighborhoods. As a consequence, optical flow-based registration does not tolerate large transformations.

Matching based on color contributions, on the other hand, shifts the emphasis to the "content" of the neighborhoods at the price of somewhat less localization accuracy. The transformation between the two image patches is taken into account through *continuous* spatial kernels and, therefore, an optimization procedure can recover large transformations. The necessary condition is weaker, only an overlap between *the two neighborhoods* containing significant color information should exist. In our approach, both methods employ the same parametrization for the underlying transformation and, thus, they can be combined into a single estimation process using the optimization criterion

$$\mathcal{J} = \mathcal{J}_{cd} + \mathcal{J}_{of}, \tag{25}$$

where the two parts were defined in (7) and (19). Since the different kernels and loss functions introduce normalizations, there is no need to weigh the two parts in (25). The minimization procedure remains the same.

The same example used in the two previous sections is employed for a comparative performance evaluation of the joint estimation procedure. Fig. 6 shows the pixelwise difference between the aligned image patches for all three procedures. The failure of the optical flow based registration for significant transformations is illustrated in Fig. 6a where the pixels have large errors. The relative low localization accuracy of color distribution matching is illustrated in Fig. 6b, where a systematic error of about two pixels is present. Only the joint minimization has small registration errors (Fig. 6c).

Success of the minimization is contingent upon an initial overlap between the two image patches. The standard technique to fulfill this condition is to use a multiresolution representation [4], [65]. First, the correspondence is established at a low resolution and is then propagated down to the original image. Our optimization process starts at the third level of an isotropic Gaussian pyramid, generated with 9×9 smoothing filters (11).

Point Correspondence Algorithm

1. Find salient point features in the reference image with the Harris corner detector.
2. Build three levels of a Gaussian pyramid for both images.
3. Define a coarse uniform grid (35 pixel steps) on the top level of the second image's pyramid. Starting from each grid point, run the optimization procedure. Select the result yielding the smallest matching score, i.e., the total residual error.
4. Refine the parameter estimate by propagating it down in the hierarchy.

Note that the algorithm *searches* for the point correspondences in the second image and it does not require putative matches. The color distribution matching component eases the requirements for an initial overlap between the image patches and a much coarser grid can be used than with the exclusively optical flow-based methods [4], [65].

2.5 Illumination Compensation

The matching process becomes more difficult when illumination changes exist between the two images. Methods based on color histograms/distribution are especially affected by the presence of color illumination differences.

Approximative illumination invariance of imaged objects is achieved in [12] by using color channel normalization and, in [15], by deriving a single invariant color coordinate as a linear combination of log RGB values. We introduce an additional parameter in the estimation process to represent the multiplicative coefficient of the color illumination change. Thus, instead of using normalized color channels, the relative luminance information is preserved at the expense of an additional parameter to estimate.

Using the illumination compensation parameter λ, the optical flow error (5) becomes

$$\boldsymbol{e}_{of}(\boldsymbol{x}_i, \boldsymbol{\theta}, \lambda) = \lambda \boldsymbol{I}_2[\boldsymbol{m}(\boldsymbol{x}_i, \boldsymbol{\theta})] - \boldsymbol{I}_1(\boldsymbol{x}_i), \tag{26}$$

while the color distribution error (16) is for a given kernel orientation β

$$e_{cd}(u, \beta, \boldsymbol{\theta}, \lambda) = \sqrt{q_{u\beta}(\boldsymbol{\theta}, \lambda)} - \sqrt{p_{u\beta}}, \tag{27}$$

where

$$q_{u\beta}(\boldsymbol{\theta}, \lambda) = C_q \sum_{j=1}^{n_m} K_{e\beta}\left(\frac{\boldsymbol{x}(\boldsymbol{m}_j, \boldsymbol{\theta}) - \boldsymbol{x}_0}{h}\right) K_e\left(\frac{\lambda \boldsymbol{c}_j - \boldsymbol{c}_u}{b}\right). \tag{28}$$

The estimation procedure remains the same, just extended with the additional parameter λ.

The advantage of introducing the illumination compensation parameter is illustrated in Fig. 7. Three interest points with their neighborhoods are shown in Fig. 7a. The corresponding points in the second image found without using the illumination compensation are shown in Fig. 7b and by using the illumination compensation in Fig. 7c. Their neighborhoods are deformed according to the estimated local transformation. By mapping the image patches from the first image into the second image, the alignment errors are shown in Figs. 7d and 7e, respectively. Clearly, by using illumination compensation (Fig. 7e), the alignment is better and the three points are correctly matched. A second example (Fig. 8) shows the three neighborhoods in the first image (Fig. 8a) and their mapping into the second image without (Fig. 8b) and with (Fig. 8c) illumination compensation. Note the incorrect alignment in Fig. 8b. The two examples show that the mapped neighborhoods have similar intensity in the second image, proving the effectiveness of illumination compensation through a multiplicative factor.

3 Performance Evaluation

To test the performance of the point matching algorithm under controlled conditions, an 820×632 image was

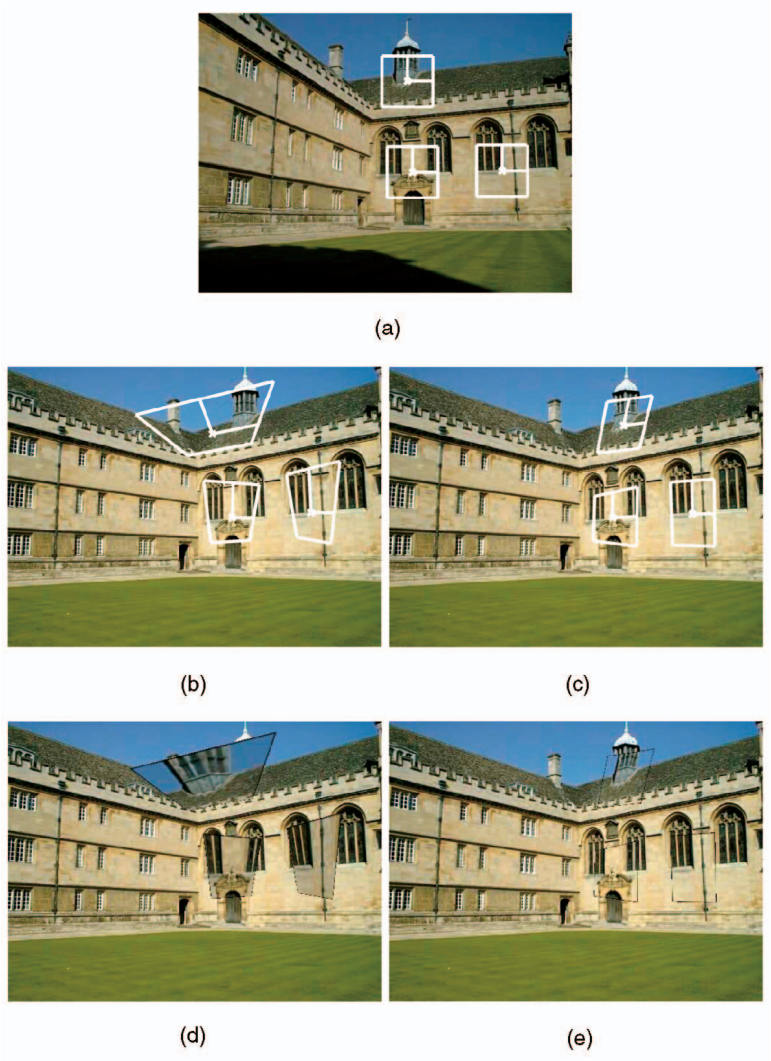

Fig. 7. *Illumination compensation experiment*. (a) The reference image and the neighborhoods of three interest points. The transformed neighborhood in the second image found without (b) and with (c) illumination compensation. Mapping of the three reference neighborhoods into the second image without (d) and with (e) illumination compensation.

Fig. 8. *Illumination compensation experiment, Valbonne*. (a) The reference image and the neighborhoods of three interest points. The mapped neighborhoods into the second image without (b) and with (c) illumination compensation.

deformed with known homographies introducing an increasing amount of distortion. After the transformation the image was resampled with the original grid (Fig. 9a).

Correspondences for 121 salient points from the original image were sought. Since ground truth was available, the *rate of detection* (Fig. 9b) and the *localization accuracy* (Fig. 9c)

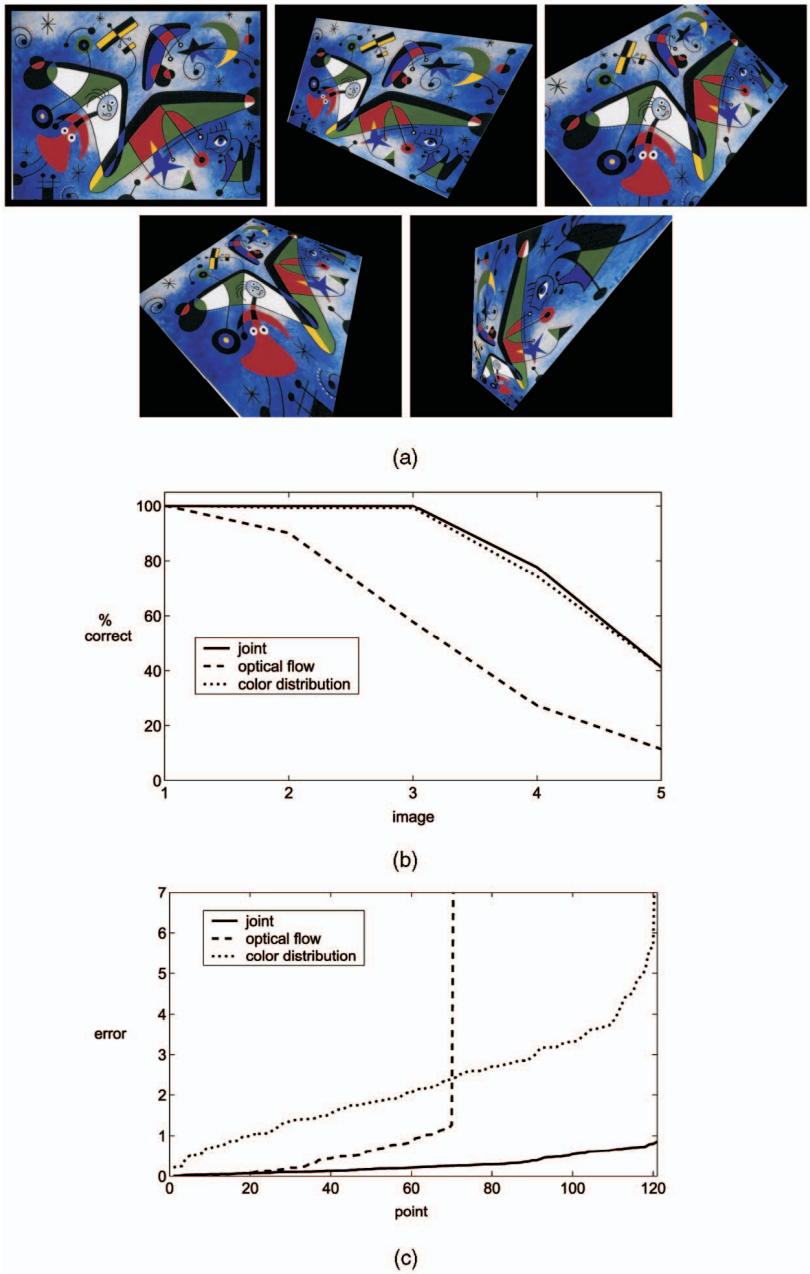

Fig. 9. *Quantitative assessment of the matching performance*. (a) Test images. The image at the top left was distorted by four known homographies. (b) The detection rate and (c) the localization accuracy are computed for the top-right image in (a).

were measured in the third image (Fig. 9a top-right). The detection rate was defined as a "hit" in the second image within 5 pixels of the ground truth. Three different experimental conditions were investigated:

1. only optical flow-based registration,
2. only color distribution matching, and
3. combined method.

The results clearly reveal the advantage of the joint approach. While the detection rate of the color distribution matching and the combined method are not distinguishable, they exhibit very different localization accuracy. The matches found by the former can have up to a seven pixel distance from the ground truth. The combined method locates *all* the correspondences within one pixel distance (Fig. 9c). The color distribution matching succeeds to bring the two neighborhoods close enough that the optical flow can become effective. Once the neighborhoods have a significant overlap, the color distribution score does not vary too much. The optical flow alone cannot cope with the large distortions in the images and fails to locate about half of the correspondences. Sequential application of the two components, color distribution-based matching followed by optical flow, is not optimal since as can be seen in Fig. 9c; the former may not bring the two image patches close enough that the latter is guaranteed to succeed.

In the subsequent experiments, the combined method was applied to several images also used in [39]. The results are shown in Fig. 10. In each case, salient points were detected in the first image (top), and their correspondences were located by the algorithm in the second image (bottom).

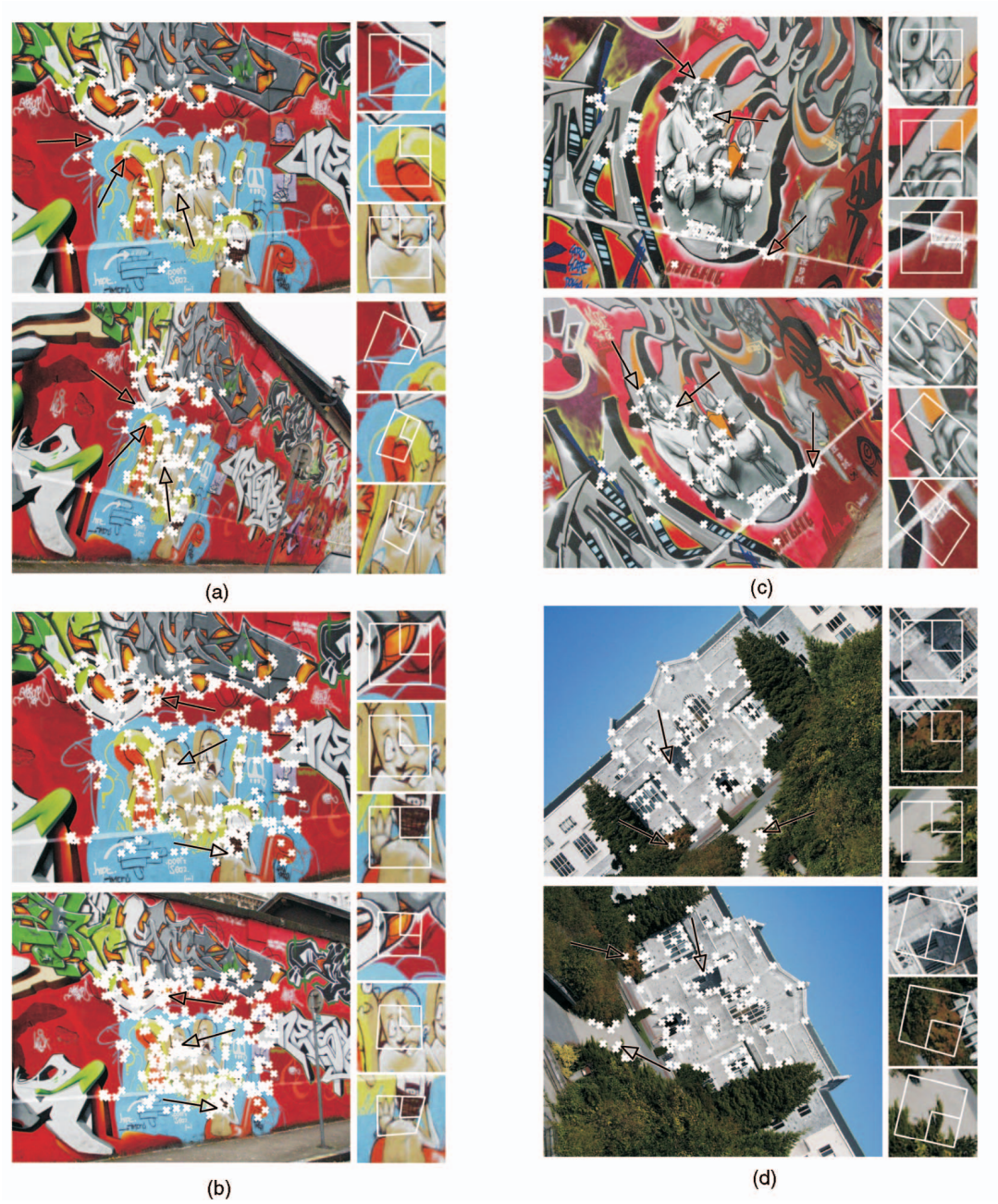

Fig. 10. *Four matching experiments*. The interest points shown in the top image were matched in the bottom image. The neighborhoods of three points (marked with arrows) are enlarged to show the estimated local transformation.

The neighborhoods associated with three points (marked by arrows) are shown on the right.

All images were 800×600 and significant scale, rotation and projective distortions are present. To measure the quality of the point matches, a *global* homography between the two images was robustly computed based only on the correspondences with a better than median matching score. *All* the established matches were then classified as inliers or outliers relative to this homography with the value of the decision threshold being 5 pixels.

In Fig. 10a, out of 128 feature points from the first image, 114 were declared as correctly matched in the second image. In Fig. 10b, the threshold of the Harris corner detector was lowered and out of 416 feature points, 308 were correctly matched. For the images in Figs. 10c and 10d, the correct detection rates were 81 out of 146 and 109 out of 137, respectively.

We can conclude that in spite of large deformations between the two images, there are enough correct correspondences that a robust global estimation technique, such as RANSAC [16] or MLESAC [59], can reliably recover 3D information. This can be exploited in wide-baseline stereo applications, e.g., [47], [50].

4 Oriented Tracking of Objects

The point matching algorithm described in this paper can be applied to tracking moving objects in an image sequence. Motion analysis, methods, the underlying models, and the supporting assumptions are reviewed in [41]. Our approach

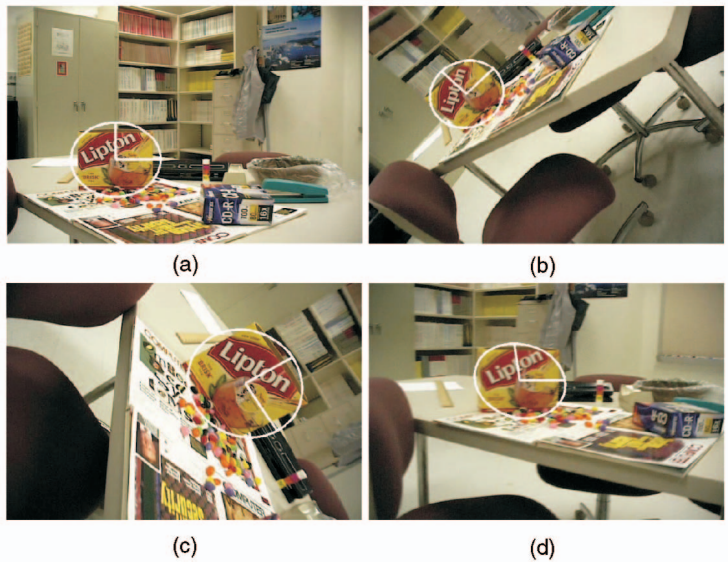

Fig. 11. *Oriented tracking of a rigid object*. (a) Initial frame. (b) Frame 73. (c) Frame 180. (d) Frame 259. (All trademarks remain the property of their respective owners. All trademarks and registered trademarks are used strictly for educational and scholarly purposes and without intent to infringe on the mark owners.)

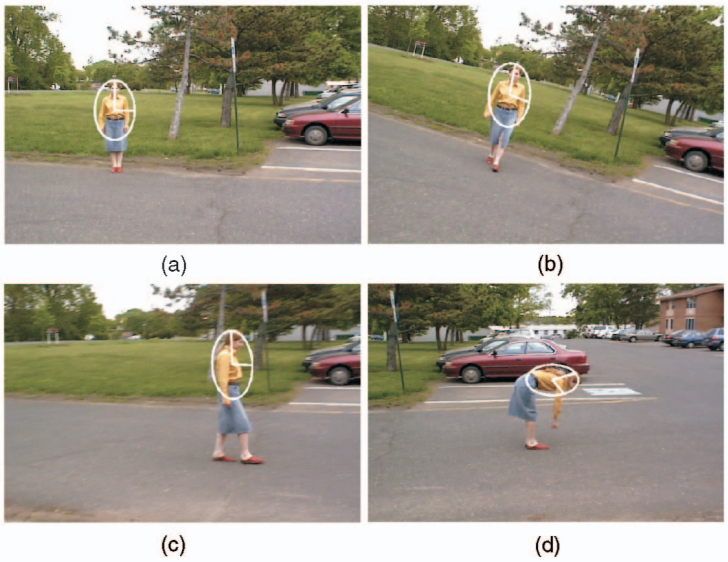

Fig. 12. *Oriented tracking of a person*. (a) Initial frame. (b) Frame 74. (c) Frame 253. (d) Frame 417.

combines the standard differential tracker [52] with the mean shift-based tracker [11]. Three improvements relative to the tracking method presented in [11] are obtained: the relative orientation of the object in each frame is also determined; the localization accuracy is much higher and the scale parameter is now automatically estimated from the data instead of being updated through explicit search. By the nature of the tracking task, once the target is defined, in the next frame, a reliable initial value for θ is usually available, therefore, the point correspondence algorithm is implemented *only* at the resolution of the input. Since, once a point correspondence is established the neighborhood of interest is also localized, tracking is automatically achieved.

In Fig. 11, four frames of a 319 frame image sequence captured by a hand held camera are shown. The camera movement included large 3D rotations and translations. The initial neighborhood of the target region was selected manually (Fig. 11a). The target model is continuously updated as in [11]. In spite of the large variations in the appearance of the target, the object of interest was successfully tracked across the entire sequence (Figs. 11b to 11d). Another example is illustrated in Fig. 12 in four frames from a 472 frame sequence which shows the oriented tracking of a person. Because of the nonrigid transformations, only the color distribution with an affine motion model was used. The person was successfully tracked in spite of the large variations in appearance.

5 Estimation Methods and Covariances

Parameter estimation using the least squares method assumes that only the values of the color vectors are affected by noise. That is, the noise does not affect the spatial gradient but only the temporal derivative. However, this is not true as it has been repeatedly discussed in the optical flow context in [42], [43], [8] and [28, p. 369]. Because of finite difference approximation of the gradient, the noise

corrupting the spatial gradient is point dependent [8]. Using least squares in the presence of point dependent noise yields biased estimates. It has also been argued that the bias of the optical flow estimate is related to human visual perception [14]. Due to the bias, the velocity estimates tend to be smaller in length and closer to the dominant gradient direction in the considered image patch.

Beside the traditional least squares, other techniques based on the errors-in-variables (EIV) model can also be used. For point dependent (heteroscedastic) noise, several general techniques have been proposed such as the HEIV method [36], the renormalization method [28] and the FNS method [10]. For estimation in the context of optical flow, see [8], [42], [43].

To illustrate the difference between the estimation techniques, using the notations from Appendix B, we rewrite the linear constraint $z_i = \boldsymbol{y}_i^\top \boldsymbol{\theta}$, $i = 1 \dots n$ as $\boldsymbol{b}_i^\top \boldsymbol{\omega} = 0$, $\boldsymbol{b}_i = [\boldsymbol{y}_i \ -z_i]^\top$, $\boldsymbol{\omega} = [\boldsymbol{\theta} \ 1]^\top$. The criterion to be minimized in the presence of heteroscedastic noise is

$$\mathcal{J}_{HEIV} = \sum_{i=1}^{n} \frac{(\boldsymbol{b}_i^\top \boldsymbol{\omega})^2}{\boldsymbol{\omega}^\top \mathrm{C}_{\mathrm{b}_i} \boldsymbol{\omega}}, \tag{29}$$

where C_{b_i} is the point dependent covariance matrix. The parameter estimate is obtained by iteratively solving a generalized eigenproblem [28], [36].

Without considering the point dependent noise process, i.e., all $\mathrm{C}_{b_i} = \mathrm{I}$, the total least squares (TLS) estimator is obtained.

$$\mathcal{J}_{TLS} = \sum_{i=1}^{n} \left(\frac{\boldsymbol{b}_i^\top \boldsymbol{\omega}}{\|\boldsymbol{\omega}\|} \right)^2 \tag{30}$$

having the solution the eigenvector corresponding to the smallest eigenvalue of the moment matrix $\mathrm{B} = \sum_{i=1}^{n} \boldsymbol{b}_i \boldsymbol{b}_i^\top$. TLS minimizes the sum of orthogonal distances from the data/measurements to the hyperplane having the normal $\boldsymbol{\omega}$. By using a simplified noise process, the TLS estimate is biased [8]. For optical flow computation, it has been also observed that the TLS estimate has larger variability than LS [43], [60], and alternative solutions have been proposed such as constrained TLS (CTLS) [60] at the expense of higher computational cost. We have used the least squares approach in this paper since the bias is not significant enough to corrupt the matching, as our experimental results have shown. Least squares is also faster and is the most stable approach (due to the presence of the small bias). Our algorithm, however, can be easily adapted to use the HEIV method either for the entire estimation process or only as a postprocessing step.

Feature detection and matching is only a preprocessing module in many geometric computation algorithms. In subsequent processing it is often of interest to use information about the location uncertainty of the matched points. The most straightforward method is to compute the covariance matrix of the point features directly from the image either through a residual-based approach or a derivative-based approach [44], [30]. In the residual approach, a quadratic function is fitted to the sum-of-squared-distances (SSD) correlation surface and the Hessian yields the inverse of the normalized covariance matrix. In the derivative approach, the first order approximation of the Hessian is given by the weighted sum of the image gradients with respect to each of the coordinates.

In our approach, the uncertainty of the parameter estimate is a byproduct of the optimization algorithm. For the least squares technique, the covariance matrix of the parameters $\hat{\mathrm{C}}_\theta$ is

$$\hat{\mathrm{C}}_\theta = \sigma^2 \sum_{i=1}^{n} \hat{\boldsymbol{y}}_i \hat{\boldsymbol{y}}_i^\top \qquad \sigma^2 = \frac{\sum_{i=1}^{n} (z_i - \hat{\boldsymbol{y}}_i^\top \boldsymbol{\theta})^2}{n - r}, \tag{31}$$

where r is the rank of the problem and $\hat{\boldsymbol{y}}_i$ are the corrected measurements.

Several studies investigated the influence of the derived point location uncertainty in the subsequent geometric computation algorithms [30], [9]. To see the difference between the covariance matrices obtained with our method and the ones obtained directly from the image gray levels, we used both of them for estimating the epipolar geometry with the unbiased HEIV estimator which takes into account the different covariances associated with each data point. For details about this estimator, see [37].

Sixteen points were selected in the first image (Fig. 13a) and matched in the second (Fig. 13b). For the second image, the estimated covariance matrices using the gray levels of the color image converted to gray scale are shown in Fig. 13c and using our method in Fig. 13d. There is a scale ambiguity of the gray scale computed covariances, however, a common global scale does not influence the fundamental matrix estimate. Thus, it suffices to compare only the shapes of the covariance matrices. It can be observed that the shapes are different when computed with the two methods.

The fundamental matrix was then computed with both sets of covariance matrices using the HEIV method while imposing the ancillary constraints [36] (p. 130). To obtain the "ground truth" estimate F_0, the HEIV method was applied to 35 points carefully selected manually. The difference in Frobenius norm between the estimated fundamental matrices and the "ground truth" is small but slightly larger when gray scale based covariance matrices are used

$$\begin{aligned} &\|\mathrm{F}_{match} - \mathrm{F}_0\| = 0.0009960 \quad \text{and} \\ &\|\mathrm{F}_{gray} - \mathrm{F}_0\| = 0.0013766. \end{aligned} \tag{32}$$

Since the location estimates in the second image have high accuracy, very good estimates are obtained for the fundamental matrix in both cases. While it was reported in [30] that the gain of using feature point covariance matrices in estimation is small, one must be aware that these covariances depend on the way they were derived.

6 Discussion

The point correspondence algorithm described in this paper is general and versatile enough to be useful in many computer vision applications. It combines two different approaches, optical flow registration and color distribution matching, in a unified robust estimation framework. The sensitivity of color distributions to the image transformation was increased by the use of oriented elliptical spatial kernels. Illumination changes are taken into account by explicit parametrization of the intensity shift and solving for this parameter in the same estimation procedure.

The proposed algorithm has superior performance, with excellent detection rate and localization accuracy for both synthetic data and real image pairs with large transformation between them. The same procedure was applied to

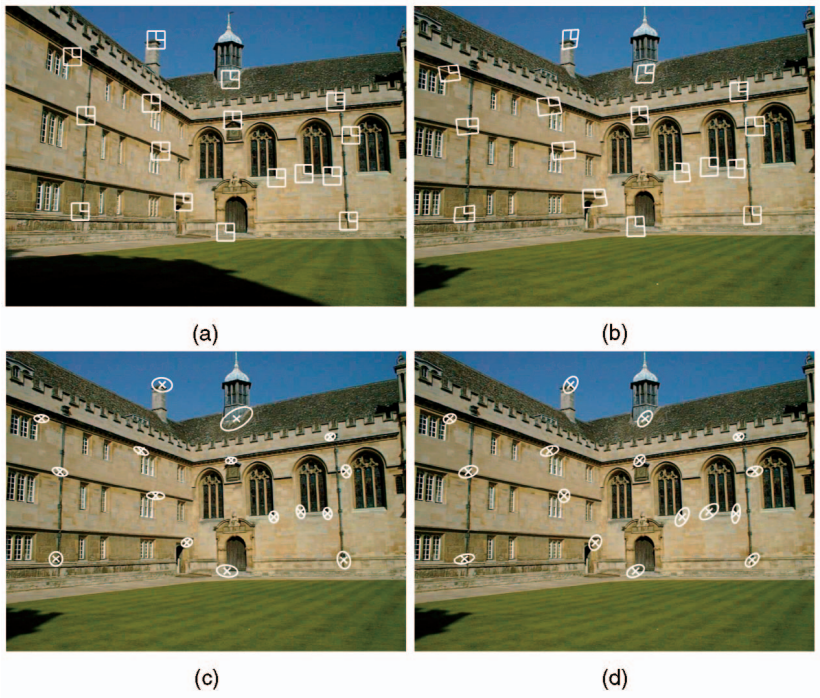

Fig. 13. *Covariance computation experiment*. The points marked in (a) were matched in (b). The shape of the covariance matrices obtained using the gray levels is shown in (c) and obtained from the point correspondence algorithm in (d).

tracking rigid and nonrigid objects in long image sequences while also recovering the relative orientation of the object.

The advantage of optical flow methodology is that of high localization accuracy, especially if the image deformation fits the transformation model, but it fails for large or nonrigid deformations. On the other hand, color distribution matching does not have high localization accuracy, but is more tolerant to large or nonrigid deformations. Combining the two through a robust framework benefits from the advantages of both methods. The procedure is flexible enough to permit activation/deactivation of the model parameters according to the desired behavior and one can use only the distribution matching as shown in the tracking application.

The performance evaluation examples in Fig. 10 were also used in [39]. The point correspondence algorithm employed there, however, is different from ours. The size of the neighborhood and its location are determined by a search in the discretized space of the parameters and affine invariant features are used for matching. We also failed to process [39], Fig. 6b, and our failure could be traced to the lack of enough consistent gray level variations in the neighborhoods of the interest points.

In case of very large scale changes between the two images, our method may not perform satisfactorily since the estimation process can try to compensate for the scale change by introducing a "false" projective transformation. For such scale changes, we recommend to run the algorithm with the initial scale set to a few representative values, and select the best result based on the matching score. Note that the correct scale will be also determined.

The proposed method works using local information independently, therefore, a possible further enhancement is to integrate the algorithm into a match validation procedure exploiting global constraints. The geometry of the projective camera provides several such constraints which can be used to restrict the motion model. Problems that arise from the presence of repetitive structures can be also avoided by global processing [62].

Our technique can be extended to other image features such as edges or filter responses along the approach proposed in [20]. Other enhancements include the use of the value of the optimization criterion (residual error) to monitor the quality of the matches [18], [29]. In a more general context, self consistency [31] can be employed to assess the adequacy of the proposed correspondence algorithm for a given computer vision task.

We have shown that combining optical flow and local distribution matching significantly improves the performance of a fundamental module in computer vision: point matching. The increase in the reliability of this module can help to build more autonomous computer vision systems. The C++ source code of the point correspondence algorithm described in the paper is available at www.caip.rutgers.edu/riul.

Appendix A
Kernel Density Estimation

This Appendix reviews kernel density (Parzen window) estimation employed in the computation of color distributions. For a more in-depth analysis of this topic, see [13], [17], [53], [63].

Given n points $\boldsymbol{x}_i$, $i = 1 \ldots n$ in a d-dimensional space $\boldsymbol{x}_i \in \mathbb{R}^d$, the multivariate kernel density estimate $\hat{p}(\boldsymbol{x})$ obtained with the kernel $K(\boldsymbol{x})$ and a symmetric positive definite bandwidth matrix H is computed at the point $\boldsymbol{x}$ as

$$\hat{p}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathrm{H}}(\boldsymbol{x}-\boldsymbol{x}_i), \tag{A.1}$$

where

$$K_{\mathrm{H}} = |\mathrm{H}|^{-1/2} K(\mathrm{H}^{-1/2}\boldsymbol{x}). \tag{A.2}$$

For the estimate $\hat{p}(\boldsymbol{x})$ to be a proper probability density function, i.e., to be nonnegative and integrate to one, the d-variate kernel $K(\boldsymbol{x})$ must satisfy the conditions

$$K(\boldsymbol{x}) \geq 0 \qquad \int_{\mathbb{R}^d} K(\boldsymbol{x})d\boldsymbol{x} = 1. \tag{A.3}$$

The accuracy of the density estimate is measured by the mean integrated squared error (MSIE), that is, the expected value of the squared error between the density estimate $\hat{p}$ and the true value p, integrated over the domain of definition. In practice, however, only an asymptotic approximation of this error measure can be computed (AMISE). Under this asymptotics, as the number of points n increase toward infinity, the elements of the full rank bandwidth matrix H should go toward zero at a rate slower than n^{-1}. For radially symmetric kernels, the AMISE criterion is minimized by the Epanechnikov kernel having the expression

$$K_e(\boldsymbol{x}) = \begin{cases} C_e(1-\boldsymbol{x}^\top\boldsymbol{x}) & \boldsymbol{x}^\top\boldsymbol{x} \leq 1 \\ 0 & \text{otherwise,} \end{cases} \tag{A.4}$$

where C_e is the normalization constant such that (A.4) is satisfied.

If the bandwidth matrix H is chosen proportional to the identity matrix $\mathrm{H} = h^2\mathrm{I}_\mathrm{d}$, then the kernel density estimate using the Epanechnikov radially symmetric kernel K_e has the expression

$$\hat{p}(\boldsymbol{x}) = \frac{1}{nh^d}\sum_{i=1}^{n} K_e\Big(\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\Big). \tag{A.5}$$

The value of the bandwidth h has to be provided and influences the degree of smoothing of the density estimate.

Selecting in two dimensions the bandwidth matrix as $\mathrm{H} = h^2\mathrm{B}_\beta$, where

$$\begin{aligned}\mathrm{B}_\beta &= \mathrm{R}_\beta \mathrm{diag}(\sigma_1^2, \sigma_2^2)\mathrm{R}_\beta^\top \\ &= \begin{bmatrix} cos\beta & -sin\beta \\ sin\beta & cos\beta \end{bmatrix}\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\begin{bmatrix} cos\beta & sin\beta \\ -sin\beta & cos\beta \end{bmatrix}\end{aligned} \tag{A.6}$$

is equivalent to employing a two-dimensional Epanechnikov elliptically symmetric kernel oriented at angle β

$$K_{e\beta}(\boldsymbol{x}) = \begin{cases} C_\mathrm{B}\Big(1-\boldsymbol{x}^\top\mathrm{B}_\beta^{-1}\boldsymbol{x}\Big) & \boldsymbol{x}^\top\mathrm{B}_\beta^{-1}\boldsymbol{x} \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{A.7}$$

and the density estimate becomes

$$\hat{p}(\boldsymbol{x}) = \frac{1}{nh^2\sigma_1\sigma_2}\sum_{i=1}^{n} K_{e\beta}\Big(\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\Big). \tag{A.8}$$

APPENDIX B
M-ESTIMATION

This Appendix reviews the family of robust estimators known as M-estimators [32], [26]. In traditional regression, a set of n data points $\boldsymbol{y}_{io}$ in $\mathbb{R}^d$ are characterized by a single parameter vector $\boldsymbol{\theta}$ through a linear relation to the true value of the measurements z_{io}

$$z_{io} = \boldsymbol{y}_{io}^\top\boldsymbol{\theta}. \tag{B.1}$$

Considering that $z_i = z_{io} + \delta z_i$ is measured with error, the parameter vector $\boldsymbol{\theta}$ is found by through minimization of the error defined with a *loss function* $\rho(u)$

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}} \sum_{i=1}^{n} \rho\Big(\frac{\boldsymbol{y}_{io}^\top\boldsymbol{\theta} - z_i}{\sigma}\Big) = \underset{\boldsymbol{\theta}}{\mathrm{argmin}}\, \mathcal{J}(\boldsymbol{\theta}), \tag{B.2}$$

where σ is the scale which controls the magnitude of the error mapped through the loss function. The loss function $\rho(u)$ should satisfy the following conditions

$$\rho(u) \geq 0 \quad \rho(0) = 0 \quad \rho(u) = \rho(-u) \quad \text{nondecreasing with } |u| \tag{B.3}$$

and should have piecewise continuous first two derivatives. Note that we can interpret the M-estimators as a generalization of the least squares estimators. By taking the loss function to be $\rho(u) = u^2$, the minimization (B.2) defaults into the well-known least squares criterion where the scale parameter σ does not influence the estimate.

The M-estimators having a loss function with bounded increase are called redescending M-estimators. In this class, we will use the *biweight* redescending loss function defined as

$$\rho(u) = \begin{cases} \frac{1}{6}\Big[1-(1-u^2)^3\Big] & |u| \leq 1 \\ \frac{1}{6} & |u| > 1. \end{cases} \tag{B.4}$$

The parameter estimate $\hat{\boldsymbol{\theta}}$ is the solution of the equation

$$\nabla_\theta \mathcal{J}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \frac{\partial\rho(u_i)}{\partial u}\frac{\partial u_i(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} = \mathbf{0}, \tag{B.5}$$

where $u_i = (\boldsymbol{y}_{io}^\top\boldsymbol{\theta} - z_i)/\sigma$. By defining the data dependent weights as

$$w_i = w(u_i) = \frac{1}{u_i}\frac{\partial\rho(u_i)}{\partial u} \tag{B.6}$$

and computing the second factor in (B.5), we obtain

$$\sum_{i=1}^{n} w_i\boldsymbol{y}_{io}(\boldsymbol{y}_{io}^\top\hat{\boldsymbol{\theta}} - z_i) = \mathbf{0}. \tag{B.7}$$

This normal equation has the weighted least squares solution

$$\hat{\boldsymbol{\theta}} = \Big(\sum_{i=1}^{n} w_i\boldsymbol{y}_{io}\boldsymbol{y}_{io}^\top\Big)^{+}\Big(\sum_{i=1}^{n} w_i\boldsymbol{y}_{io}z_i\Big), \tag{B.8}$$

where "+" represents the pseudoinverse operator. For the biweight loss function the expression of the weights derived from (B.4) is

$$w(u) = \begin{cases} (1-u^2)^2 & |u| \leq 1 \\ 0 & |u| > 1, \end{cases} \tag{B.9}$$

i.e., a redescending loss function completely removes the influence of measurements that yield large errors. Solving for the parameters estimate $\hat{\boldsymbol{\theta}}$ is done iteratively. Starting from an initial value $\hat{\boldsymbol{\theta}}^0$, the weights are derived from the error and the estimate is updated using (B.8). For the

multivariate case, the argument of the loss function is passed through its norm $\rho(\|\boldsymbol{u}\|)$.

An extension of the classical M-estimation procedure is to generalize the weights such that the influence of outliers in the y-space is bounded. These are measurements with large $\boldsymbol{y}_i$ relative to the majority of the data points, but can have z_i values similar to them. In the *generalized* M-estimation procedure, the weights are not a direct function of the error, but instead are defined as $w_i = w(z_i, \boldsymbol{y}_{io}, \boldsymbol{\theta}, \sigma)$ [32]. The conditions on the function w (nonnegative, bounded and continuous) assure that the influence of outliers in the y-space is reduced. The simplest way is to decompose this function as a product between a weight defined on the residual error and a weight defined on $\boldsymbol{y}$ [49] (p. 13), and this is the approach adopted in the paper.

APPENDIX C

JACOBIAN COMPUTATION

Following [7], the Jacobian matrix for a vector value function $\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^n$ in the variable $\boldsymbol{x} \in \mathbb{R}^m$ is the $m \times n$ matrix

$$\mathrm{J}_{f|x} \stackrel{\triangle}{=} \frac{\partial \boldsymbol{f}(\boldsymbol{x})^\top}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_m} & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix} = \left(\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top} \right)^\top. \tag{C.1}$$

For a scalar value function $f(\boldsymbol{x})$, the Jacobian becomes the gradient with respect to $\boldsymbol{x}$, $\mathrm{J}_{f|x} = \boldsymbol{g}_f$. The chain rule is used to compute the Jacobian of the composite function $\boldsymbol{f}(\boldsymbol{x}(\boldsymbol{y})), \boldsymbol{y} \in \mathbb{R}^s$, that is

$$\mathrm{J}_{f|y} = \mathrm{J}_{x|y} \mathrm{J}_{f|x}. \tag{C.2}$$

As an example, we will derive some of the elements of the Jacobian matrix of the transformation defined by (1). The equivalent relation between the point coordinates $\boldsymbol{x}$ and $\boldsymbol{m}$ is

$$\boldsymbol{m} = \frac{\mathrm{A}(\boldsymbol{x} - \boldsymbol{x}_0)}{\boldsymbol{v}^\top (\boldsymbol{x} - \boldsymbol{x}_0) + 1} + \boldsymbol{m}_0. \tag{C.3}$$

Thus, the derivative with respect to the angle α_0 parametrizing the rotation matrix R_0 is

$$\frac{\partial \boldsymbol{m}}{\partial \alpha_0} = \begin{bmatrix} -sin\alpha_0 & -cos\alpha_0 \\ cos\alpha_0 & -sin\alpha_0 \end{bmatrix} \mathrm{R}_1^\top \mathrm{S} \mathrm{R}_1 \frac{(\boldsymbol{x} - \boldsymbol{x}_0)}{\boldsymbol{v}^\top (\boldsymbol{x} - \boldsymbol{x}_0) + 1} \tag{C.4}$$

and the derivative with respect to the projective parameters $\boldsymbol{v}^\top$ is

$$\frac{\partial \boldsymbol{m}}{\partial \boldsymbol{v}^\top} = -\frac{\mathrm{A}(\boldsymbol{x} - \boldsymbol{x}_0)}{\left[\boldsymbol{v}^\top (\boldsymbol{x} - \boldsymbol{x}_0) + 1\right]^2} (\boldsymbol{x} - \boldsymbol{x}_0)^\top. \tag{C.5}$$

The remaining values, as well as the Jacobian of the inverse transform, are derived the same way.

ACKNOWLEDGMENTS

The support of the NSF grant IRI 99-87695 is gratefully acknowledged. The authors would like to thank Krystian Mikolajczyk and Cordelia Schmid for providing some of the images used in the experiments. This work was done while Dr. Georgescu was with the Computer Science Department, Rutgers University, 94 Brett Road, Piscataway, NJ 08854-8058.

REFERENCES

[1] S. Baker and I. Matthews, "Equivalence and Efficiency of Image Alignment Algorithms," *Proc. 2001 IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 1090-1097, Dec. 2001.

[2] J. Barron and R. Klette, "Quantitative Color Optical Flow," *Proc. 16th Int'l Conf. Pattern Recognition,* vol. 4, pp. 251-255, Aug. 2002.

[3] A.C. Berg and J. Malik, "Geometric Blur for Template Matching," *Proc. 2001 IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 607-614, Dec. 2001.

[4] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani, "Hierarchical Model-Based Motion Estimation," *Proc. European Conf. Computer Vision,* pp. 237-252, May 1992.

[5] M.J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parameteric and Piecewise-Smooth Flow Fields," *Proc. Conf. Computer Vision and Image Understanding,* vol. 63, pp. 75-104, 1996.

[6] J.W. Brandt, "Improved Accuracy in Gradient-Based Optical Flow Estimation," *Int'l J. Computer Vision,* vol. 25, pp. 5-22, 1997.

[7] J.W. Brewer, "Kronecker Products and Matrix Calculus in System Theory," *IEEE Trans. Circuits and Systems,* vol. 25, pp. 772-781, 1978.

[8] J. Bride and P. Meer, "Registration via Direct Methods: A Statistical Approach," *Proc. 2001 IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 984-989, Dec. 2001.

[9] M.J. Brooks, W. Chojnacki, D. Gawley, and A. van den Hengel, "What Value Covariance Information in Estimating Vision Parameters?" *Proc. Eighth Int'l Conf. Computer Vision,* vol. 1, pp. 302-308, July 2001.

[10] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley, "From FNS to HEIV: A Link between Two Vision Parameter Estimation Methods," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, pp. 264-268, 2004.

[11] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 142-149, June 2000.

[12] M.S. Drew, J. Wei, and Z.N. Li, "Illumination-Invariant Image Retrieval and Video Segmentation," *Pattern Recognition,* vol. 32, pp. 1369-1388, 1999.

[13] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification,* second ed. Wiley, 2000.

[14] C. Fermuller, D. Shulman, and Y. Aloimonos, "The Statistics of Optical Flow," *Computer Vision and Image Understanding,* vol. 82, pp. 1-32, Apr. 2001.

[15] G.D. Finlayson and S.D. Hordley, "Color Constancy at a Pixel," *J. Optical Soc. of Am. A,* vol. 18, pp. 253-264, 2001.

[16] M.A. Fischler and R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. Assoc. Computing Machinery,* vol. 24, pp. 381-395, 1981.

[17] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* second ed., Academic Press, 1990.

[18] A. Fusiello, E. Trucco, T. Tommasini, and V. Roberto, "Improving Feature Tracking with Robust Statistics," *Pattern Analysis and Applications,* vol. 2, pp. 312-320, 1999.

[19] T. Gevers and A.W.M. Smeulders, "Content-Based Image Retrieval by Viewpoint-Invariant Color Indexing," *Image and Vision Computing,* vol. 17, pp. 475-488, 1999.

[20] V. Govindu and C. Shekhar, "Alignment Using Distributions of Local Geometric Properties," *IEEE Trans. Pattern Analysis Machine Intelligence,* vol. 21, pp. 1031-1043, 1999.

[21] A. Griffin and J. Kittler, "An Active Mesh Based Tracker for Improved Feature Correspondences," *Pattern Recognition Letters,* vol. 23, pp. 443-449, 2002.

[22] A.B. Hadiashar and D. Suter, "Robust Optic Flow Computation," *Int'l J. Computer Vision,* vol. 29, pp. 59-77, 1998.

[23] G.D. Hager and P.N. Belhumeur, "Real-Time Tracking of Image Regions with Changes in Geometry and Illumination," *Proc. 1996 IEEE Conf. Computer Vision and Pattern Recognition,* pp. 403-410, June 1996.

[24] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Alvey Vision Conf.,* pp. 147-151, 1988.

[25] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, 2000.

[26] P.J. Huber, *Robust Statistical Procedures,* second ed. Soc. Industrial and Applied Math., 1996.

[27] M. Irani and P. Anandan, "Robust Multi-Sensor Image Alignment," *Proc. Fourth Int'l Conf. Computer Vision,* pp. 959-966, Jan. 1998.

[28] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice.* Elsevier, 1996.
[29] K. Kanatani and Y. Kanazawa, "Automatic Thresholding for Correspondence Detection," *Proc. Statistical Methods in Video Processing Workshop,* pp. 19-24, June 2002.
[30] Y. Kanazawa and K. Kanatani, "Do We Really Have to Consider Covariance Matrices for Image Features?" *Proc. Eighth Int'l Conf. Computer Vision,* vol. 2, pp. 301-306, July 2001.
[31] Y.G. Leclerc, Q.T. Luong, and P. Fua, "Self-Consistency and MDL: A Paradigm for Evaluating Point Correspondence Algorithms and Detecting Change," *Proc. Int'l J. Computer Vision,* vol. 51, pp. 63-83, 2003.
[32] G. Li, "Robust Regression," *Exploring Data Tables, Trends, and Shapes,* pp. 281-343, D.C. Hoaglin, F. Mosteller, and J.W. Tukey, eds., John Wiley and Sons, 1985.
[33] D.G. Lowe, Object Recognition from Local Scale-Invariant Features," *Proc. Seventh Int'l Conf. Computer Vision,* pp. 1150-1157, Sept. 1999.
[34] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with Application to Stereo Vision," *Proc. Int'l Joint Conf. Artificial Intelligence,* pp. 674-679, Aug. 1981.
[35] J. Matas, S. Obdrzalek, and O. Chum, "Local Affine Frames for Wide-Baseline Stereo," *Proc. 16th Int'l Conf. Pattern Recognition,* vol. 4, pp. 363-366, Aug. 2002.
[36] B. Matei, "Heteroscedastic Errors-In-Variables Models in Computer Vision," PhD thesis, Dept. of Electrical and Computer Eng., Rutgers Univ., 2001. Available at http://www.caip.rutgers.edu/riul/research/theses.html.
[37] B. Matei and P. Meer, "A General Method for Errors-in-Variables Problems in Computer Vision," *Proc. 2000 IEEE Computer Vision and Pattern Recognition Conf.,* vol. 2, pp. 18-25, June 2000.
[38] P. Meer and I. Weiss, "Smoothed Differentiation Filters for Images," *J. Visual Comm. and Image Representation,* vol. 3, pp. 58-72, 1992.
[39] K. Mikolajczyk and C. Schmid, "An Affine Invariant Interest Point Detector," *Proc. European Conf. Computer Vision,* vol. 1, pp. 128-142, May 2002.
[40] F. Mindru, T. Moons, and L.V. Gool, "Comparing Intensity Transformations and Their Invariants in the Context of Color Pattern Recognition," *Proc. European Conf. Computer Vision,* vol. 4, pp. 448-460, May 2002.
[41] A. Mitiche and P. Bouthemy, "Computation and Analysis of Image Motion: A Synopsis of Current Problems and Methods," *Proc. Int'l J. Computer Vision,* vol. 19, pp. 29-55, 1996.
[42] H. Nagel, "Optical Flow Estimation and the Interaction between Measurement Errors at Adjacent Pixel Positions," *Proc. Int'l J. Computer Vision,* vol. 15, pp. 271-288, 1995.
[43] L. Ng and V. Solo, "Errors-in-Variables Modeling in Optical Flow Estimation," *IEEE Trans. Image Processing,* vol. 10, pp. 1528-1540, 2001.
[44] K. Nickels and S. Hutchinson, "Estimating Uncertainty in SSD-Based Feature Tracking," *Image and Vision Computing,* vol. 20, pp. 47-58, 2002.
[45] M. Nicolescu and G. Medioni, "Perceptual Grouping from Motion Cues Using Tensor Voting in 4-D," *Proc. European Conf. Computer Vision,* vol. 3, pp. 423-437, May 2002.
[46] J.M. Odobez and P. Bouthemy, "Robust Multiresolution Estimation of Parametric Motion Models Applied to Complex Scenes," *J. Visual Comm. and Image Representation,* vol. 6, pp. 348-365, 1995.
[47] P. Pritchett and A. Zisserman, "Wide Baseline Stereo Matching" *Proc. Sixth Int'l Conf. Computer Vision,* pp. 754-760, Jan. 1998.
[48] F. Rothganger, S. Lazebnik, C. Schmidt, and J. Ponce, "3D Object Modeling and Recognition Using Affine-Invariant Patches and Multi-View Spatial Constraints," *Proc. 2003 IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 272-280, June 2003.
[49] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection.* Wiley, 1987.
[50] F. Schaffalitzky and A. Zisserman, "Multi-View Matching for Unordered Image Sets, or How Do I Organize My Holiday Snaps?" *Proc. European Conf. Computer Vision,* vol. 1, pp. 414-431, May 2002.
[51] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of Interest Point Detectors," *Computer Vision and Image Understanding,* vol. 78, pp. 151-172, 2000.
[52] J. Shi and C. Tomasi, "Good Features to Track," *Proc. 1994 IEEE Conf. Computer Vision and Pattern Recognition,* pp. 593-600, June 1994.
[53] B.W. Silverman, *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 1986.
[54] R. Szeliski and J Coughlan, "Spline-Based Image Registration," *Proc. Int'l J. Computer Vision,* vol. 22, pp. 199-218, 1997.
[55] D. Tell and S. Carlsson, "Wide Baseline Point Matching Using Affine Invariants Computed from Intensity Profiles," *Proc. European Conf. Computer Vision,* pp. 814-828, June 2000.
[56] D. Tell and S. Carlsson, "Combining Appearance and Topology for Wide Baseline Matching," *Proc. European Conf. Computer Vision,* vol. 1, pp. 68-81, May 2002.
[57] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto, "Making Good Features Track Better," *Proc. 1998 IEEE Conf. Computer Vision and Pattern Recognition,* pp. 178-183, June 1998.
[58] P.H.S. Torr and C. Davidson, "IMPSAC: Synthesis of Importance Sampling and Random Sample Consensus," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, pp. 354-364, 2003.
[59] P.H.S. Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry," *Computer Vision and Image Understanding,* vol. 78, pp. 138-156, 2000.
[60] C.J. Tsai, N.P. Galatsanos, and A.K. Katsaggelos, "Optical Flow Estimation from Noisy Data Using Differential Techniques," *Proc. 1999 IEEE Int'l Conf. Acoustics, Speech and Signal Processing,* vol. 6, pp. 3393-3396, Mar. 1999.
[61] T. Tuytelaars and L. Van Gool, "Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions," *Proc. 11th British Machine Vision Conf.,* pp. 412-425, Sept. 2000.
[62] T. Tuytelaars, A. Turina, and L. Van Gool, "Noncombinatorial Detection of Regular Repetitions under Perspective Skew," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, pp. 418-432, 2003.
[63] M.P. Wand and M.C. Jones, *Kernel Smoothing.* Chapman & Hall, 1995.
[64] B. Wang, K.K. Sung, and T.K. Ng, "The Localized Consistency Principle for Image Matching under Non-Uniform Illumination Variation and Affine Distortion," *Proc. European Conf. Computer Vision,* vol. 1, pp. 205-219, May 2002.
[65] J. Weber and J. Malik, "Robust Computation of Optical-Flow in a Multiscale Differential Framework," *Int'l J. Computer Vision,* vol. 14, pp. 67-81, 1995.

Bogdan Georgescu received the Dipl. Engn. degree in 1996, the MS degree in 1997 in electrical engineering from the Bucharest Polytechnic Institute, Bucharest, Romania, and the MS degree in 2001 in computer science from Rutgers University, Piscataway, New Jersey. He is currently a PhD student in the Department of Computer Science at Rutgers University. His research interests are in computer vision, machine learning, and statistical pattern recognition. He is a student member of IEEE and the IEEE Computer Society.

Peter Meer received the Dipl. Engn. degree from the Bucharest Polytechnic Institute, Romania in 1971 and the DSc degree from the Technion, Israel Institute of Technology, Haifa, in 1986, both in electrical engineering. From 1971 to 1979, he was with the Computer Research Institute, Cluj, Romania, working on R&D of digital hardware. From 1986 to 1990, he was an assistant research scientist at the Center for Automation Research, University of Maryland at College Park. In 1991, he joined the Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey, and is currently a professor. He has held visiting appointments in Japan, Korea, Sweden, Israel, and France, and was on the organizing committees of numerous international workshops and conferences. He was an associate editor of the *IEEE Transaction on Pattern Analysis and Machine Intelligence* between 1998 and 2002, is a member of the editorial board of *Pattern Recognition*, and was a guest editor of *Computer Vision and Image Understanding* for a special issue on "robust statistical techniques in image understanding." He is coauthor of an award winning paper in *Pattern Recognition* in 1989, the best student paper in the 1999, and the best paper in the 2000 *IEEE Conference on Computer Vision and Pattern Recognition*. His research interest is in application of modern statistical methods to image understanding problems. He is a senior member of the IEEE.