# MDL, Collineations and the Fundamental Matrix

S.J. Maybank and P.F. Sturm, Department of Computer Science, The University or Reading, Whiteknights, Reading, Berkshire, RG6 6AY, UK.

(S.J.Maybank,P.F.Sturm)@reading.ac.uk

**Abstract**

Scene geometry can be inferred from point correspondences between two images. The inference process includes the selection of a model. Four models are considered: background (or null), collineation, affine fundamental matrix and fundamental matrix. It is shown how Minimum Description Length (MDL) can be used to compare the different models. The main result is that there is little reason for preferring the fundamental matrix model over the collineation model, even when the former the 'true' model.

## 1 Introduction

Model selection is a central task in computer vision: given data obtained from images and given a number of models, which model is most strongly supported by the data? Is it better to have *i)* a simple model fitting the data approximately; or *ii)* a complicated model fitting the data very closely [1, 6, 7, 9, 10, 15, 18, 20]? Accuracy of fit to the data is by itself not a sufficient criterion for choosing a model. The fit can always be improved by allowing a greater flexibility or generality in the model. In many cases, a sufficiently general model fits the data with zero error.

In simple cases, the allowable models are specified by probability density functions $p(x|\theta)$ defined on the data $x$ and depending on a parameter vector $\theta$ with a fixed dimension. In such cases the Maximum Likelihood (ML) principle is a good, widely used strategy: given $x$, select the value of $\theta$ at which $p(x|\theta)$ attains its maximum value. The ML principle fails if the dimension of $\theta$ can vary. Tt is necessary to introduce a penalty for the number of model parameters, otherwise a model with a large number of parameters will always be chosen in preference to models with only a few parameters.

In the literature there are several suggestions for penalising overparameterisation, for example [1, 18]. The Bayesian Information Criterion (BIC) of [18] applies if the allowable models can be divided into separate families such that the ML principle holds for each family separately. The BIC yields for each family an error criterion of the form $\log(p(x|\hat{\theta})) - (1/2)D\log(N)$ where $\hat{\theta}$ is the maximum likelihood estimate of $\theta$ for the family, $D$ is the total number of parameters in the model and $N$ is the total number of measurements. If the number of measurements is low, then the second term dominates and models with low $D$ are favoured. If the number of measurements is high, then the first term dominates, because it depends on the fit of the model to the large number of measurements, and the value of $D$ is less important.

The BIC is applicable only if probability density functions are defined on the space of possible errors in the measurements and on the space of parameter values for each family

of models. These density functions are prior information which can strongly affect the model choice.

Minimum Description Length (MDL) [9, 10] is an alternative way of comparing models. Unlike Bayesian methods it does not require explicit probability density functions for the data and the parameters. In MDL the data are first expressed as a bit string $s$. If a model $\mathcal{M}$ fits the data, then $s$ contains internal structure depending on $\mathcal{M}$. This structure is removed, to give a compressed string. The compressed string and a bit string description of $\mathcal{M}$ are concatenated to give a string $U_{\mathcal{M}}(s)$. The model $\mathcal{M}$ is strongly favoured if $U_{\mathcal{M}}(s)$ is much shorter than $s$. The compression must be information preserving, in that $s$ can be recovered exactly from $U_{\mathcal{M}}(s)$. Overparameterised models are penalised simply because they require a long description.

The Shannon-Fano code [10] is an example of MDL. The code is for symbols supplied randomly and independently. If a symbol $a$ has probability $p$ then the optimal 0,1 code for $a$ has length $\lceil -\log(p) \rceil$, where log is the logarithm to base 2. The key point here is the close link between the model ($a$ supplied with probability $p$) and the code length.

Applications of MDL to computer vision can be found in [3, 5, 8, 11, 13, 14]. In many cases, for example [8, 11], MDL is used to assign a prior probability to a model. The deviations of the data from the model are given probabilities assuming a standard distribution such as the Gaussian. The probability of a model conditional on the data is derived using Bayes' theorem, and then maximised over the model parameters.

In this paper MDL is applied to sets of pairs of corresponding points obtained from images of a room taken by a digital camcorder. The aim is to investigate a test case application of MDL, and to see if it performs as expected. The models are $\mathcal{B}$, $\mathcal{C}$, $\mathcal{A}$, $\mathcal{F}$, as listed in §1.1. No explicit probabilistic assumptions are made. The only criterion for comparing the models is the length of the compressed bit strings $U_{\mathcal{B}}(s)$, $U_{\mathcal{C}}(s)$, $U_{\mathcal{A}}(s)$, $U_{\mathcal{F}}(s)$. The main new result is that $\mathcal{C}$ achieves a good compression in all cases, in particular,

$$|U_{\mathcal{F}}(s)| \approx |U_{\mathcal{C}}(s)| \text{ when the 'true' model is } \mathcal{F}. \tag{1}$$

Without more information, for example, additional images or constraints on the shapes of objects, there is little reason for ever preferring $\mathcal{F}$ to $\mathcal{C}$.

## 1.1 The models

To be specific, suppose the data are the pixel coordinates of corresponding points $q_i \leftrightarrow q_i'$, $1 \leq i \leq n$ in two images of the same scene. Two points $q$, $q'$ in different images correspond if and only if they are projections of the same scene point [4]. The images are embedded in the projective plane, $\mathbb{P}^2$, by adding 1 as a third coordinate, $(x, y) \mapsto (x, y, 1)$. There are many possible models, each of which involves assumptions about the relative position of the two cameras or the scene geometry [18, 19]. In this paper the following models are considered.

- $\mathcal{B}$) Background: the image points have no discernable structure.
- $\mathcal{C}$) Collineation: there is a collineation $\omega : \mathbb{P}^2 \to \mathbb{P}^2$ such that $\omega(q_i) = q_i'$, $1 \leq i \leq n$.
- $\mathcal{A}$) Affine fundamental: there is a $3 \times 3$ matrix $A$ with rank 2 such that $A_{11} = A_{12} = A_{21} = A_{22} = 0$ and $q_i^\top A q_i' = 0$, $1 \leq i \leq n$ [17].
- $\mathcal{F}$) Fundamental: there is a $3 \times 3$ matrix $F$ with rank 2 such that $q_i^\top F q_i' = 0$, $1 \leq i \leq n$.

### 1.2 Notation

A bit string is an element of $\{0,1\}^*$. The length of a bit string $s$ is $|s|$. The floor function is $x \mapsto \lfloor x \rfloor$ and the ceiling function is $x \mapsto \lceil x \rceil$, where $\lfloor x \rfloor$ is the greatest integer such that $\lfloor x \rfloor \leq x$, and $\lceil x \rceil$ is the least integer such that $x \leq \lceil x \rceil$.

The fixed length code of length $b$ for a non-negative integer $n$ is $d(n,b)$. The usual binary code for $n$ is padded with zeros on the right to give a bit string of length $b$. If $n \geq 1$, then $d(n,b)$ is defined only if $b \geq \lceil \log(n) \rceil + 1$, where log is the logarithm to base 2.

The logstar prefix code $r : \mathbb{N}^+ \to \{0,1\}^*$ is implemented as described in [2], §4.3.2. Note that $|r(n)| \approx \log^*(n) + 1.51857\ldots$. A general discussion of the logstar code, including the definition of $\log^*$, can be found in [10], with further information in [2]. The map zton : $Z \to \mathbb{N}^+$ is $\text{zton}(n) = 2n$ for $n \geq 1$, and $\text{zton}(n) = 2\,\text{abs}(n) + 1$ for $n < 1$, and the logstar code is redefined on $Z$ by $e(n) = r(\text{zton}(n))$.

## 2 Coding the Data

In this section the strings $U_{\mathcal{B}}(s)$, $U_{\mathcal{C}}(s)$, $U_{\mathcal{A}}(s)$, $U_{\mathcal{F}}(s)$ are defined. All the codes are constructed using rational arithmetic, in order to avoid inaccuracies arising from floating point approximations. The algorithms are implemented in Mathematica [21].

The image points are defined for $1 \leq i \leq n$ by $q_i = (x_i, y_i, 1)$, $q_i' = (x_i', y_i', 1)$ where $x_i$, $y_i$, $x_i'$, $y_i'$ are integers. If the feature points are located in each image to an accuracy of 1 pixel then the $x_i$, $y_i$ etc. are the pixel coordinates. If feature points are located with subpixel accuracy, then the $x_i$, $y_i$ etc. are scaled pixel coordinates. The code $c : Z^n \to \{0,1\}^*$ used in this section is defined below in §3.

### 2.1 Background $\mathcal{B}$

Let $x, y, x', y' \in Z^n$ be the vectors with respective components $x_i$, $y_i$, $x_i'$, $y_i'$. The code for $q_i \leftrightarrow q_i'$, $1 \leq i \leq n$ under the background model $\mathcal{B}$ is $U_{\mathcal{B}}(s) = c(x).c(y).c(x').c(y')$.

### 2.2 Collineation $\mathcal{C}$

Let $\omega : \mathbb{P}^2 \to \mathbb{P}^2$ be a collineation, ie. a map of the the form $q \mapsto Hq$, where $H$ is an invertible $3 \times 3$ matrix, and let $a_i$, $b_i$, $\epsilon_i$, $\delta_i$ be defined for $1 \leq i \leq n$ by

$$\begin{aligned} \omega(q_i) &= (a_i, b_i, 1)^\top \\ (\epsilon_i, \delta_i) &= (\lfloor x_i' - a_i + 0.5 \rfloor, \lfloor y_i' - b_i + 0.5 \rfloor) \end{aligned}$$

The point $q_i'$ can be recovered exactly from $\omega$, $q_i$ and the integers $\epsilon_i$, $\delta_i$.

Let $\epsilon, \delta \in Z^n$ be the vectors with respective coordinates $\epsilon_i$, $\delta_i$ and let code($\omega$) be a coding of $\omega$. The $q_i$, $q_i'$, $1 \leq i \leq n$ are coded by the string

$$t = c(x).c(y).\text{code}(\omega).c(\epsilon).c(\delta) \tag{2}$$

If $\omega$ is a good fit to the data, then $\epsilon$, $\delta$ are small and compression is achieved.

How should code($\omega$) be constructed? A key issue is the precision with which the components of $\omega$ are specified. If the precision is low, then $|\text{code}(\omega)|$ is small but $|c(\epsilon).c(\delta)|$ is

large. If the precision is high, then $|\text{code}(\omega)|$ is large but $|c(\epsilon).c(\delta)|$ is small. The problem of choosing the best precision is circumvented by using RANSAC [16].

Let $u_i \leftrightarrow v_i$, $1 \leq i \leq 4$ be pairs of corresponding points in $\mathbb{P}^2$, such that no three of the $u_i$ are collinear and no three of the $v_i$ are collinear. There is a unique collineation $\omega$ such that $\omega(u_i) = v_i$, $1 \leq i \leq 4$. A coding of the $u_i$, $v_i$, $1 \leq i \leq 4$ yields a coding of $\omega$. Ideally, all quadruples $q_{i_j} \leftrightarrow q'_{i_j}$, $1 \leq j \leq 4$ of corresponding points should be examined to find the quadruple for which

$$|\text{code}(\omega).c(\epsilon).c(\delta)| \tag{3}$$

is a minimum. In practice there are too many quadruples, so a random selection of $N$ quadruples is made and (3) is minimised over the chosen quadruples.

An advantage of RANSAC is that the precision of $\omega$ is appropriate for the data; in addition, the code for $\omega$ is redundant because it includes the points $q_{i_j}$, $1 \leq j \leq 4$ already coded in $c(x).c(y)$. The redundancy is removed and compression achieved by omitting the $q_{i_j}$ from code($\omega$), and instead coding the index of the four-tuple $\iota = (i_1, i_2, i_3, i_4)$, $i_1 < i_2 < i_3 < i_4$ in the list of all ordered four-tuples with distinct entries drawn from $n$. The code length for $\iota$ is at most $\lceil \log(b(n,4)) \rceil + 1$ bits where $b(n,4)$ is the binomial coefficient.

Further compression of $t$ in (2) is achieved by omitting from $\epsilon$, $\delta$ the eight entries known to be zero, yielding the code $U_{\mathcal{C}}(s)$.

## 2.3 Affine fundamental matrix $A$

Let $A$ be an affine fundamental matrix, and let $l$ be the line $l' = q^\top A$. The geometrical interpretation of the equation $q^\top A q' = 0$ is that $q'$ lies on $l'$. If $q$, $A$ are given, then $q'$ can be coded by giving its position on $l'$. Compression is achieved because only one coordinate is needed rather than two.

As with $\mathcal{C}$, RANSAC is used to find a suitable matrix $A$ compatible with the $q_i \leftrightarrow q'_i$, $1 \leq i \leq n$. Let $u_i \leftrightarrow v_i$, $1 \leq i \leq 4$ be pairs of corresponding points in $\mathbb{P}^2$ such that no three of the $u_i$ are collinear, no three of the $v_i$ are collinear, none of the $u_i$, $v_i$ are on the line at infinity and there is no affine transformation T such that $Tu_i = v_i$, $1 \leq i \leq 4$. Then there is a unique affine fundamental matrix $A$ such that $u_i^\top A v_i = 0$, $1 \leq i \leq 4$.

The point $q'_i$ is specified relative to an origin which depends on $i$, the $q_j$ and $A$. In detail, there is a three dimensional family of collineations which preserve the epipolar lines associated with $A$ in that if $\rho$ is any one of the collineations and $l$ is an epipolar line of $A$ in the first image, then $\rho(l)$ is the corresponding epipolar line in the second image [12]. The three dimensional family is spanned the collineations associated with any four linearly independent matrices $H$ for which

$$AH + H^\top A^\top = 0 \tag{4}$$

Let $q_{i_j} \leftrightarrow q'_{i_j}$, $1 \leq j \leq 4$ be the pairs of corresponding points which define $A$. From the $q_{i_j}$ select the three points $q_{i_j}$, $q_{i_k}$, $q_{i_l}$ which define a triangle with the greatest area. A unique collineation is specified by the matrix $H$ for which $Hq_{i_j} = q'_{i_j}$, $Hq_{i_k} = q'_{i_k}$, $Hq_{i_l} = q'_{i_l}$ and (4) holds.

Let $\nu_i$ be a unit vector in $\mathbb{R}^2$ parallel to $q_i^\top A$, let $\nu_i^\perp$ be a unit vector perpendicular to $\nu_i$, and define $r_i$, $s_i$ by $q'_i - Hq_i = (r_i \nu_i + s_i \nu_i^\perp, 1)$ as shown in Figure 1. Define integers

$\epsilon_i$, $\delta_i$ by

$$(\epsilon_i, \delta_i) = (\lfloor 2r_i + 0.5 \rfloor, \lfloor 2s_i + 0.5 \rfloor) \tag{5}$$

The factor 2 on the right hand side of (5) is needed to remove quantisation errors.

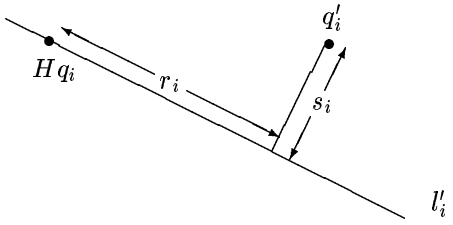


Figure 1. Definition of $r_i$, $s_i$.

The code for the $q_i \leftrightarrow q_i'$, $1 \leq i \leq n$ is

$$c(x).c(y).\text{code}(A).c(\epsilon).c(\delta) \tag{6}$$

The matrix $A$ is specified by giving the index of $\iota = (i_1, i_2, i_3, i_4)$ in the list of ordered four-tuples of distinct elements drawn from $n$. When coding $\delta$, the four entries known to be zero are omitted.

A random selection of $N$ quadruples is made and the length of the code (6) minimised over the quadruples. The code with the minimum length is $U_{\mathcal{A}}(s)$.

In a full reconstruction of the 3D scene the collineation $H$ described above defines a plane in space which passes near to the 3D points projecting down to corresponding points in the two images.

## 2.4 Fundamental matrix $\mathcal{F}$

The coding of $s$ as $U_{\mathcal{F}}(s)$ is similar to the coding as $U_{\mathcal{A}}(s)$, with one significant change, due to the fact that four pairs of image correspondences are not sufficient to specify a unique fundamental matrix. Let $q_{i_j} \leftrightarrow q'_{i_j}$, $1 \leq j \leq 7$ be seven pairs of corresponding points. There are in general exactly two linearly independent $3 \times 3$ matrices $F_1$, $F_2$ such that $q_{i_j}^\top F q'_{i_j} = 0$, $1 \leq j \leq 7$. The fundamental matrices compatible with the $q_{i_j} \leftrightarrow q'_{i_j}$, $1 \leq i \leq 7$ are obtained by solving the cubic polynomial equation in $t$ [19],

$$\det(F_1 + tF_2) = 0 \tag{7}$$

There are at most three real roots. To specify a unique fundamental matrix it is necessary to record the appropriate root of (7), which requires two bits.

Let $t_j$ be a real root of (7) and let $\tilde{F} = F_1 + t_j F_2$. The matrix $\tilde{F}$ is replaced by a rational approximation $F$, retaining the constraint $\det(F) = 0$. Let $\tilde{u}$ be the eigenvector of $\tilde{F}^\top \tilde{F}$ with the least eigenvalue, let $u$ be a rational approximation to $\tilde{u}$ and let $G$ be a rational approximation to $\tilde{F}$. The matrix $F$ is defined by

$$F = G - (u.u)^{-1} Gu \otimes u$$

# 3 Code for Vectors in $Z^n$

If $c_1, \ldots, c_p$ are different codes for vectors $x \in Z^n$, then a new code $c$ can be constructed by first finding the index $j$ such that

$$|c_j(x)| = \min\{|c_i(x)|, 1 \leq i \leq p\}$$

and then setting $c(x) = d(j,b).c_j(x)$, where $d$ is defined in §1.4. If $p$ is small and $n$ is large, then $c$ may give shorter average code lengths than any single code $c_i$.

The code $c$ in §2 is constructed from four separate codes $c_1$, $c_2$, $c_3$, $c_4$, which are described in turn.

## 3.1 Codes $c_1$ and $c_2$

Let $C_\sigma$ be defined for $\sigma \in \mathbb{N}$ by

$$C_\sigma = \{x \in Z^n, |x_i| \leq \sigma, 1 \leq i \leq n\}$$

The set $C_\sigma$ contains $(2\sigma+1)^n$ points. The elements of $C_\sigma$ are enumerated by a function $\zeta_\sigma : C_\sigma \to \mathbb{N}^+$ constructed such that $\zeta_0(0) = 1$, and such that $\zeta_\sigma$ is an extension of $\zeta_{\sigma-1}$ for $\sigma \geq 1$. The functions $\zeta_\sigma$, $\sigma \geq 0$ together define a single function $\zeta : Z^n \to \mathbb{N}^+$.

Let $m$ be the median of the $x_i$ and let $v$ be the vector with components $v_i = x_i - m$, $1 \leq i \leq n$. The codes $c_1$, $c_2$ are defined by

$$c_1(x) = r(\zeta(x)) \qquad \text{and} \qquad c_2(x) = r(\text{zton}(m)).c_1(v)$$

## 3.2 Code $c_3$

Let $\sigma = \text{abs}(x_j)$ for some $j$, let $u$ be the vector of components $x_i$ such that $\text{abs}(x_i) \leq \sigma$, and let $v$ be the vector of components $x_i - \text{sign}(x_i)\sigma$ for those $i$ such that $|x_i| > \sigma$. Let $w_\sigma \in \{0,1\}^n$ be defined such that $(w_\sigma)_i = 1$ if $x_i$ is a component of $u$ and $(w_\sigma)_i = 0$ if $x_i - \text{sign}(x_i)\sigma$ is a component of $v$.

The code $d_\sigma$ is defined by $d_\sigma(x) = w_\sigma.c_1(u).c_1(v)$. Let $\theta$ be the value of $\sigma$ at which $|d_\sigma(x)|$ is a minimum over all the distinct values of $\sigma = \text{abs}(x_i)$, $1 \leq i \leq n$, that is $|d_\theta(x)| = \min_\sigma\{|d_\sigma(x)|\}$. The code $c_3$ is defined by $c_3(x) = d_\theta(x)$.

## 3.3 Code $c_4$

Let $m$ be the median of $x$, let $y_1, \ldots, y_p$ be the distinct integers appearing in the set $\{x_i - m, 1 \leq i \leq n\}$, and let $y_i$ occur $k_i$ times, $1 \leq i \leq p$. Let $S$ be the set of all permutations of $x$. The number $b$ of elements of $S$ is given by the multinomial $b = n!/(k_1! \ldots k_m!)$. The elements of $S$ are ordered in any convenient way. Let $\iota$ be the index of $x$ in the chosen order, let $y$, $k$ be the vectors with the respective components $y_i$, $k_i$, and let $m_k$ be median of $k$. The code $c_4$ is defined by

$$\begin{aligned} f(x) &= e(k_1 - m_k).e(y_1). \ \ldots \ .e(k_p - m_k).e(y_p) \\ c_4(x) &= e(m).e(m_k).f(x).d(\iota, b) \end{aligned}$$

If the $x_i$ are independent realisations of a random variable and $n$ is large, then $|c_4(x)|/n$ is, with a high probability, close to the shortest possible expected length of a code word.

See for example [10], §1.11.4. In practice the effectiveness of $c_4$ is reduced because of the extra code needed for $y$, $k$.

Figure 2. Images for collineation model.

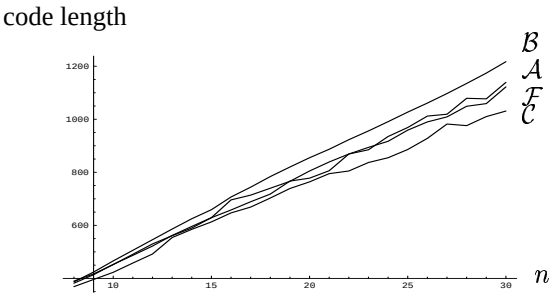


Figure 3. Code lengths when the 'true' model is a collineation

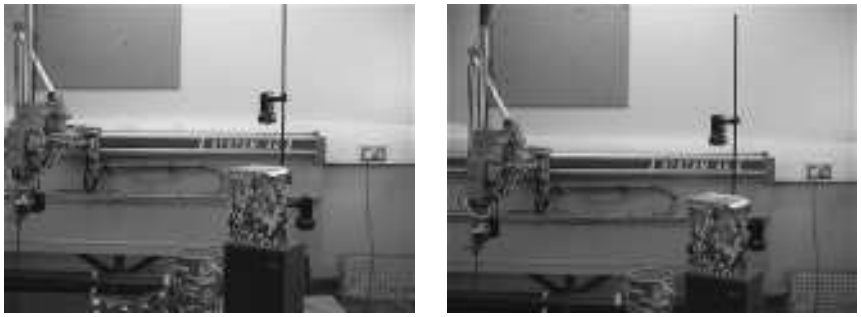

Figure 4. Images for affine fundamental matrix model.

# 4 Experiments

Images of a laboratory were taken by a Canon MV-1 Camcorder mounted on a tripod. Typical pairs of images are shown in Figures 2, 4, 6 with the 'true' models shown in the captions. In each case the 'true' model is known, but only because of the prior information available to a human observer. Figure 2 shows two images of a flat poster, Figure 4 is

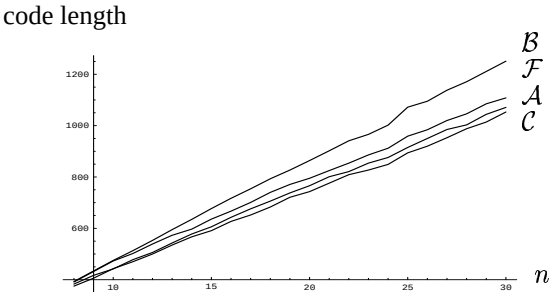


Figure 5. Code lengths when the 'true' model is an affine fundamental matrix

obtained by translating the camera parallel to the image plane, and in Figure 6 the camera is moved forwards to produce significant projective distortions of the image.

The task of the program is to make the best choice of model using only the data $q_i \leftrightarrow q_i'$, $1 \leq i \leq n$ and the models $\mathcal{B}, \mathcal{C}, \mathcal{A}, \mathcal{F}$. This best choice can and does differ from the 'true' model.

The size of the original images in pixels is $640 \times 480$. Feature points were located in each image and matched to obtain pairs of corresponding points $q_i \leftrightarrow q_i'$, $1 \leq i \leq n$. The graphs of code length against $n$ for $8 \leq n \leq 30$ are shown in Figures 3, 5, 7. The maximum number on the vertical scale is 1200 bits, and the spacing between numbers is 200 bits. The number of random samples in the RANSAC algorithm was $N = 10$. Increasing the value of $N$ as far as 20 did not produce significant changes in the graphs. Higher values were not investigated because of the long run times.

It is apparent from the graphs that $\mathcal{C}$ is always the preferred model even when the 'true' model is $\mathcal{F}$. The models $\mathcal{A}$, $\mathcal{F}$ show a similar performance, and $\mathcal{B}$ is always the worst model.

Figure 6. Images for fundamental matrix model.

# 5 Discussion

The experiments show that the collineation model $\mathcal{C}$ is a good choice even for sets of image correspondences for which the 'true' model is a fundamental matrix. This is in agreement with the comment in [18], Section 4, that for two images, simple models are strongly favoured over more complex ones. Why does the model $\mathcal{F}$ perform so badly

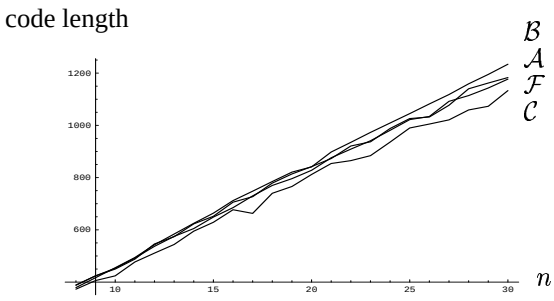


Figure 7. Code lengths when the 'true' model is a fundamental matrix

under MDL? The reason can be seen in Figure 1. In the usual methods for assessing the fit of $\mathcal{F}$ to the data, the error measure is the sum of the squares of the $s_i$, and the $r_i$, measuring distances along the epipolar lines, are ignored. In MDL the $r_i$ must be included, to obtain a loss free coding of the data. The extra code length needed for the $r_i$ reduces the preference for $\mathcal{F}$, so that in these experiments $\mathcal{C}$ is almost always preferred. This argument suggests that $\mathcal{C}$ will still be preferred if all the image pixel values are used rather than than just the locations of a few salient points.

If additional information is given, then $\mathcal{F}$ may become the preferred model. For example, suppose that $\mathcal{C}$ is augmented by an assumption that $q_i' - \omega(q_i)$ is a realisation of a Gaussian random variable with a known covariance, and that $q_i' - \omega(q_i)$ is coded under this assumption. When the 'true' model is a fundamental matrix, then the code for $q_i' - \omega(q_i)$ will be long and $\mathcal{F}$ will be preferred to $\mathcal{C}$.

Simple parameter counting, in agreement with general arguments based on the BIC, suggest that for long image sequences a generalisation of $\mathcal{F}$ will be preferred over a model in which pairs of images are related by a collineation. For example, suppose that images 1,2,3 are given with fundamental matrices $F_{12}$, $F_{23}$, $F_{13}$. Let $q \leftrightarrow q' \leftrightarrow q''$ be a triple of corresponding points. Then $q''$ is determined by $q$, $q'$ and $F_{13}$, $F_{23}$, because it is the intersection of the epipolar lines $q^\top F_{13}$ and $q'^\top F_{23}$. If the $F_{ij}$ are replaced by collineations $H_{ij}$ such that $H_{ij}$ preserves the epipolar lines associated with $F_{ij}$, then there is no certain way of locating $q''$ given only $q$, $q'$ and the $H_{ij}$. This suggests that the fundamental matrix model will yield a shorter code for the points in the third image.

# References

[1] H. Akaike, "A new look at statistical model identification," *IEEE Trans. Automation and Control*, Vol. 19, pp. 716-723, 1974.

[2] R. Baxter, "Minimum message length inductive inference - theory and applications," *Phd Thesis, Department of Computer Science, Monash University, Aus.*, 1996.

[3] T. Darrell and A. Pentland, "Cooperative robust estimation using layers of support," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 17, pp. 474-487, 1995.

[4] O.D. Faugeras, *Three-Dimensional Computer Vision: a geometric viewpoint*. MIT Press, 1993.

[5] P. Fua and A.J. Hanson, "An optimization framework for feature extraction," *Machine Vision and Applications*, Vol. 4, pp. 59-87, 1991.

[6] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier, 1996.

[7] K. Kanatani, "Comments on nonparametric segmentation of curves into various representations." *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 1391-1392, 1997.

[8] Y.G. Leclerc, "Constructing simple stable descriptions for image partitioning," *International Journal of Computer Vision*, Vol. 3, pp. 73-102, 1989.

[9] M. Li and P.M.B. Vitányi, "Inductive reasoning and Kolmogorov complexity," *J. of Computer and System Sciences*, Vol. 44, pp. 343-384, 1982.

[10] M. Li and P. M. B. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Graduate Texts in Computer Science, Springer, 2nd edition, 1997.

[11] T. Lindeberg and M.-X. Li, "Segmentation and classification of edges using minimum description length approximation and complementary junction clues," *Computer Vision and Image Understanding*, Vol. 67, pp. 88-98, 1997.

[12] Q.-T. Luong and T. Viéville, "Canonic representations for the geometries of multiple projective views," In J.-O. Eklundh (ed.) *Computer Vision-ECCV'94, Vol I*, Lecture Notes in Computer Science, Vol. 800, pp. 589-599, Springer, 1994.

[13] S.J. Maybank and R. Fraile, "Minimum description length method for facet matching," *Proc. International Symposium on Multispectral Image Processing, ISMIP'98*, SPIE Vol. 3545, Wuhan, China, pp. 330-335, 1998.

[14] A. Pentland, "Part segmentation for object recognition," *Neural Computation*, Vol. 1, pp. 82-91, 1989.

[15] P.L. Rosin and G.A.W. West, "Response to Kanatani," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 1393-1394, 1997.

[16] P.J. Rousseeuw, *Robust Regression and Outlier Detection*, John Wiley, 1987.

[17] L.S. Shapiro, A. Zisserman and M. Brady "Motion from point matches using affine epipolar geometry". In J.-O. Eklundh (ed.) *Computer Vision - ECCV'94, Vol. II*, Lecture Notes in Computer Science, Vol. 801, pp. 73-84, Springer, 1994.

[18] P. H. S. Torr and A. Zisserman, "Concerning Bayesian motion segmentation, model averaging, matching and the trifocal tensor," In H. Burkhardt and B. Neumann (eds.) *Computer Vision - ECCV'98, Vol. I*, Lecture Notes in Computer Science, Vol. 1406, pp. 511-527, Springer, 1998.

[19] P.H.S. Torr, A. Zisserman and S.J. Maybank, "Robust detection of degenerate configurations whilst estimating the fundamental matrix," *Computer Vision, Graphics, and Image Processing*, Vol. 71, pp. 312-333, 1998.

[20] C.S. Wallace and P.R. Freeman, "Estimation and inference by compact coding," *J. Royal Stat. Soc. Series B*, Vol. 49, pp. 240-265, 1987.

[21] S. Wolfram, *The Mathematica Book*, 3rd Edition, Cambridge University Press, Cambridge, 1996.