# Matching Images with Different Resolutions

Yves Dufournaud* Cordelia Schmid Radu Horaud

INRIA RHÔNE-ALPES & GRAVIR-CNRS
655 av. de l'Europe
38330 Montbonnot, France

Yves.Dufournaud@inrialpes.fr

## Abstract

*In this paper we address the problem of matching two images with two different resolutions: a high-resolution image and a low-resolution one. On the premise that changes in resolution act as a smoothing equivalent to changes in scale, a scale-space representation of the high-resolution image is produced. Hence the one-to-one classical image matching paradigm becomes one-to-many because the low-resolution image is compared with all the scale-space representations of the high-resolution one. Key to the success of such a process is the proper representation of the features to be matched in scale-space. We show how to extract interest points at variable scales and we devise a method allowing the comparison of two images at two different resolutions. The method comprises the use of photometric- and rotation-invariant descriptors, a geometric model mapping the high-resolution image onto a low-resolution image region, and an image matching strategy based on the robust estimation of this geometric model. Extensive experiments show that our matching method can be used for scale changes up to a factor 6.*

## 1 Introduction

The problem of matching two images has been an active topic of research in computer vision for the last two decades. The vast majority of existing methods consider two views of the same scene where the viewpoints differ by small offsets in position, orientation and viewing parameters such as focal length. Under such conditions, the images associated with the two views have comparative resolutions and hence they encapsulate scene features at approximatively the same scale. In this paper we address a somehow different problem that has only little been addressed in the past. We consider for matching two images with very different resolutions.

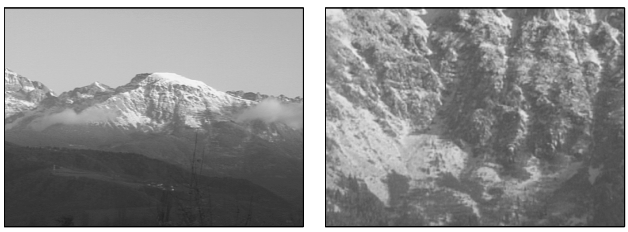

**Figure 1. An example of matching a low-resolution image with a high-resolution one.**

More precisely, if we denote by $d$ the approximate distance from an observed scene object to a viewpoint and by $f$ the focal length associated with the viewing parameters, the image resolution may be defined as $r = f/d$ or more generally as a function of $r$. Therefore we are interested in developing a matching technique which takes as input a low resolution image, image #1, and a high resolution image, image #2, such that their associated resolutions $r_1$ and $r_2$ satisfy the constraint $r_1 << r_2$. In practice it will be shown that, using the approach advocated below, it is possible to match two images such that $r_2/r_1 = 6$.

As an example we consider the image pair in Figure 1. Both images were taken with a camera placed at 11 kilometers (6.9 miles) away from the top of the mountain. For the first image (left) we used a focal length equal to 12mm while for the second one (right) we used a focal length equal to 72mm. Notice that the high-resolution image corresponds to a small region of the low-resolution one and it

* Y. Dufournaud acknowledges support from Aerospatiale.

is quite difficult to find the exact position and size of this region. Clearly, a scene object and/or texture may appear at different sizes and positions in the two images.

Therefore, the search space associated with the feature-to-feature matching of two such images is larger and more complex than the one associated with the classical stereo matching paradigm. The classical approach to image matching extracts interesting point-features from each image, matches them based on cross-correlation, computes the epipolar geometry through the robust estimation of the fundamental matrix, and establishes many other matches once this matrix is known. For a number of reasons, this method cannot be applied anymore:

1. Point-feature extraction and matching are resolution dependent processes.
2. The high-resolution image corresponds to a small region of the low-resolution one and hence the latter contains many features which do not have a match in the former.
3. It may be difficult to estimate the epipolar geometry because there is not enough depth associated with the high resolution image.

The solution suggested in this paper consists of considering a scale-space representation of the high-resolution image and of matching the low-resolution image against the scale-space description of the high-resolution one. A scale-space representation may be obtained by smoothing an image with Gaussian kernels of increasing standard deviations. Therefore, the high-resolution image will be described by a discrete set of images at various scales. On the premise that decreasing the resolution can be modeled as a smoothing equivalent to a scale change, the one-to-one image matching problem at hand becomes a one-to-many image matching problem.

In this paper we describe such a matching method. Key to its success are the following features:

- The scale-space representation of interest points together with their associated descriptors.
- A geometric model describing the mapping from the high-resolution image to the low-resolution one.
- An image-matching strategy which combines point-to-point assignments with a robust estimation of the geometric mapping.

Several authors addressed the problem of matching two images gathered from two very different points of view [3, 15, 19] but they did not consider a change in resolution. The use of scale-space in conjunction with stereo matching has been restricted to hierarchical matching: correspondences obtained at low resolution constrain the search space at higher resolutions [4, 16, 10]. Scale-space properties are thoroughly studied in [11] and the same author attempted to characterize the best scale at which an image feature should be represented [12]. A similar idea is presented in [13] to detect stable points in scale space. Our work is closely related with [5] which attempts to match two images of the same object gathered with two different zoom settings. Point-to-point correspondences are characterized in scale space by correlation traces. The method is able to recover the scale factor for which two image points are the most similar but it cannot deal with camera rotations.

Image descriptors that are invariant with respect to local affine greyvalue changes, image rotations, and image translations were studied theoretical in [9] and an efficient implementation was proposed in [17]. These descriptors are based on convolutions with Gaussian kernels and their derivatives. They are therefore consistent with scale-space representations. They are best applied to interest points and a recent study showed that the Harris corner detector is the most reliable one [18]. However, they are not scale-invariant and, in spite of good theoretical models for such invariants [11, 8], it is more judicious from a practical point of view to compute local descriptors at various scales in a discrete scale-space [17].

**Paper organization.** The remainder of this paper is organized as follows. Section 2 briefly outlines the geometric model associated with the image pair. Section 3 suggests a framework for adapting the detection of interest points to scale changes. Section 4 describes the high-resolution to low-resolution matching and section 5 presents results.

## 2 Geometric modeling

One of the key observations enabling the matching of two images at two different resolutions is that the high-resolution image corresponds to a small region of the low-resolution one. Hence, one reasonable assumption is to consider that the mapping between the high resolution image and the corresponding low-resolution region is a plane projective transformation, i.e., the scene corresponding to this region is planar. Such a homography may well be represented by a $3\times3$ homogeneous full rank matrix $\mathbf{H}$. Let $m$ be a point in the first image (low resolution) and $n$ be a point in the second image (high resolution). One can characterize a region in the low-resolution image such that the

points within this region verify:

$$\boldsymbol{m}^{in} \simeq \mathbf{H}\boldsymbol{n} \tag{1}$$

Similarly, points outside this region, say $\boldsymbol{m}^{out}$ do not verify this equation. In the general case it is quite tedious to find a parameterization of $\mathbf{H}$. Moreover, image descriptors which are invariant to such a general plane-to-plane projective transformation are difficult to compute and therefore it is difficult to properly select potential candidate points satisfying eq. (1).

We can further simplify the geometric model and consider a restricted class of homographies, namely a rotation about the optical axis, a translation, and a scale factor:

$$\boldsymbol{m}^{in} = \begin{bmatrix} h\cos\theta & -h\sin\theta & a \\ h\sin\theta & h\cos\theta & b \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{n} \tag{2}$$

Notice that the projective equality in eq. (1) is replaced by an equality and two point-to-point correspondences are sufficient to linearly estimate such a similarity transformation. In practice it will be useful to replace the 3-vectors $\boldsymbol{n}$ and $\boldsymbol{m}$ used above by 2-vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ such that:

$$\boldsymbol{n} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{x} \\ 1 \end{pmatrix} \text{ and } \boldsymbol{m} = \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}' \\ 1 \end{pmatrix}$$

With this notation, eq. (2) becomes $\boldsymbol{x}' = h\mathbf{R}\boldsymbol{x} + \boldsymbol{t}$ where $\mathbf{R}$ is the $2\times2$ rotation matrix and $\boldsymbol{t}$ is the translation associated with the image transformation.

Ideally, one would like to characterize image points by descriptors invariant to image rotation, translation and scale. Unfortunately, as already outlined, scale-invariant image descriptors are hard to compute in practice. Therefore, the matching strategy will build a discrete scale space on top of the high-resolution image thus by-passing the scale-invariance problem.

The image matching problem at hand becomes the problem of (i) extracting sets of points from the two images, $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ and $(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_M)$, (ii) properly characterizing these points such that point-to-point correspondences are allowed, and (iii) determining the largest set of such correspondences compatible with a homography between the high-resolution image and a low-resolution region.

## 3 Scale-space interest point detection

In order to match two images one has to define a measure of similarity. One possible definition is correlation. In our case, this can be written as:

$$\sum_{\boldsymbol{p}\in W} \left[I'(h\mathbf{R}(\boldsymbol{x}-\boldsymbol{p})) - I(\boldsymbol{x}-\boldsymbol{p})\right]^2$$

where $W$ is a window around $\boldsymbol{x}$. Therefore, one must find a scale factor $h$ and a rotation matrix $\mathbf{R}$ for which the expression above is minimized. The search space associated with such a technique is very large and the associated non-linear minimization procedure has problems.

Alternatively, one may use interest points which are detected by a rotation-invariant operator and characterize these points by rotation-invariant descriptors.

Such an interest point detector was proposed in [6]. More precisely, consider an image point $\boldsymbol{x}$ and the associated image greyvalue $I(\boldsymbol{x})$. Interest points are detected by:

1. Compute the image derivatives in the $u$ and $v$ directions, $I_u$, and $I_v$. These computations are carried out by convolution with the differential of a Gaussian kernel of standard deviation $\sigma$.

2. Form the auto-correlation matrix. This matrix $\mathbf{C}(\boldsymbol{x}, \sigma, \tilde{\sigma})$ averages derivatives in a window around a point $\boldsymbol{x}$. A Gaussian $G(\tilde{\sigma})$ is used for weighting:

$$\mathbf{C}(\boldsymbol{x}, \sigma, \tilde{\sigma}) = G(\tilde{\sigma}) \star \begin{bmatrix} I_u^2(\boldsymbol{x}, \sigma) & I_u I_v(\boldsymbol{x}, \sigma) \\ I_u I_v(\boldsymbol{x}, \sigma) & I_v^2(\boldsymbol{x}, \sigma) \end{bmatrix} \tag{3}$$

3. $\boldsymbol{x}$ is an interest point if the matrix $\mathbf{C}$ has two significant eigenvalues, that is if the determinant and trace of this matrix verify:

$$\det(\mathbf{C}) - \alpha \,\mathrm{trace}^2(\mathbf{C}) > t \tag{4}$$

where $t$ is a fixed threshold and $\alpha$ a parameter.

Notice that the interest point detector defined above is rotation-invariant – this is due to the symmetry of matrix $\mathbf{C}$. However, IT IS NOT invariant to a change in the image size $h$ or image resolution. Without loss of generality we can therefore omit the image-plane rotation at interest points detected by the operator described above. Under the assumption that the greyvalues are properly normalized, the similarity condition that must be satisfied is $I'(\boldsymbol{x}') = I(\boldsymbol{x})$ where, as before, $I$ is the high-resolution image and $I'$ is the low resolution one. Since the rotation is omitted we have $\boldsymbol{x}' = h\boldsymbol{x} + \boldsymbol{t}$. Taking the derivatives of the above expression with respect to the image coordinates $u$ and $v$, we obtain $hI'_u = I_u$ and $hI'_v = I_v$.

Therefore, the relationship between the interest point detector applied to the high-resolution image and the interest point detector applied to the low-resolution image is:

$$\mathbf{C}'(\boldsymbol{x}', h\sigma, h\tilde{\sigma}) = \frac{1}{h^2}\mathbf{C}(\boldsymbol{x}, \sigma, \tilde{\sigma})$$

We consider now the scale-space associated with the high resolution image. The scale-space is obtained by convolving the initial image with a Gaussian kernel who's standard deviation is increasing monotonically, say $s\sigma$ with $s > 1$. At some scale $s$ in this space the high resolution image $I$ is given by:

$$I(\boldsymbol{x}, s\sigma) = I(\boldsymbol{x}) \star G(\boldsymbol{x}, s\sigma)$$

At this scale, the image's first order derivatives write:

$$I_u(\boldsymbol{x}, s\sigma) = I(\boldsymbol{x}) \star G_u(\boldsymbol{x}, s\sigma)$$
$$I_v(\boldsymbol{x}, s\sigma) = I(\boldsymbol{x}) \star G_v(\boldsymbol{x}, s\sigma)$$

Therefore, one can detect interest points at any scale $s$ by simply replacing $\sigma$ with $s\sigma$ in eqs. (3) and (4). If the task consists of matching the high-resolution image $I$ with the low-resolution one $I'$, it is crucial to select the scale of $I$ at which this matching has to be performed. The scale $s$ must "absorb" the size ratio $h$, therefore one may write $h = \frac{1}{s}$. The interest point detector at scale $s$ is defined by:

$$\mathbf{C}(\boldsymbol{x}, s\sigma, s\tilde{\sigma}) = s^2 G(s\tilde{\sigma}) \star \begin{bmatrix} I_u^2(\boldsymbol{x}, s\sigma) & I_u I_v(\boldsymbol{x}, s\sigma) \\ I_u I_v(\boldsymbol{x}, s\sigma) & I_v^2(\boldsymbol{x}, s\sigma) \end{bmatrix} \tag{5}$$

In order to illustrate the results obtained with this scale-space interest point detector, we applied it to the high-resolution image of Figure 1 (right) at 4 scales, i.e. 1,3,5 and 8. Figure 2 shows these results where $\sigma = 1$ and $\tilde{\sigma} = 2$.

## 4 Robust image matching

The scale-space extraction and representation of interest points that are rotation-invariant will enable us to devise the one-to-many image matching technique described below. The main idea is to compare the low-resolution image at one scale with the high-resolution image at many scales. Hence, the scale at which this matching process provides the best results, provides the correct one-to-one assignments between interest points. Because the matching is supported by the robust estimation of a homography between the two images, the estimated parameters will provide – among others – the resolution ratio between the two images.

Without loss of generality, we assume that the high-resolution image, image #2, is represented at 8 different scales $\sigma$, $2\sigma$, ..., $8\sigma$ with $\sigma = 1$. At each scale $s_i$, interest points are extracted using eq. (5). Furthermore, a number of differential invariants are extracted at each scale as well. These descriptors are photometric-, image rotation-, and image translation-invariant. Likewise, interest points and their descriptors are computed and associated with the low-resolution image, image #1, at only one scale, $\sigma$.

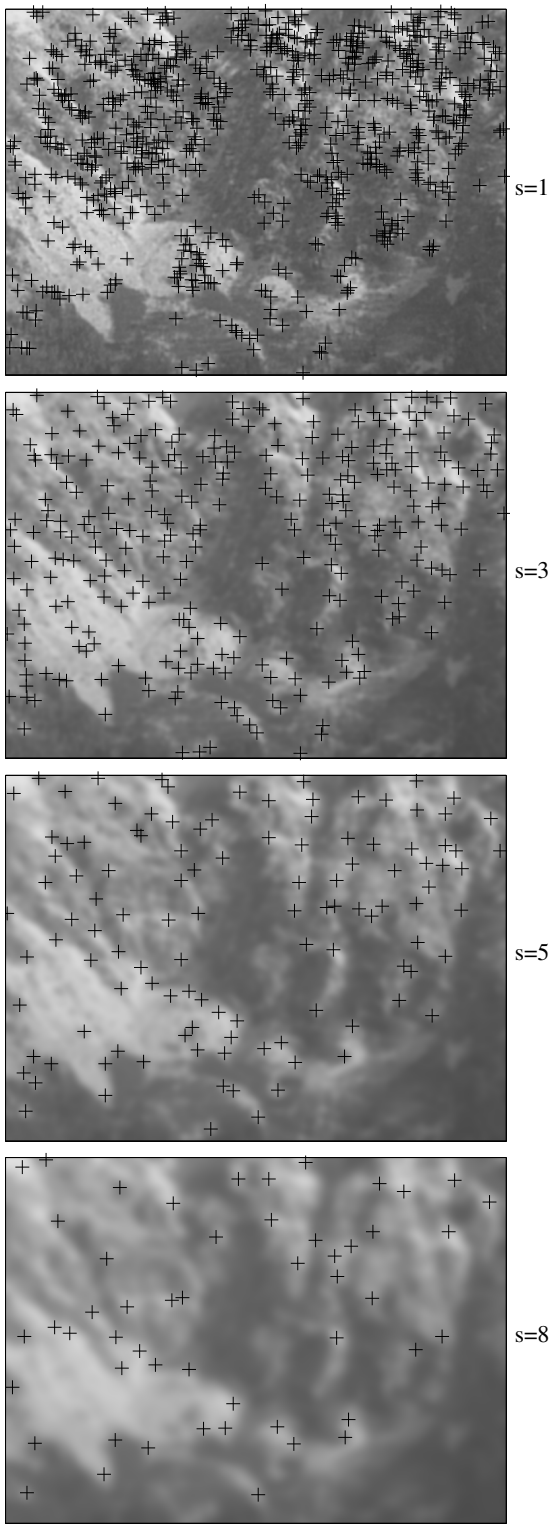


**Figure 2. Interest points detected at 4 scales.**

We then consider one-by-one the scale-space representations of image #2 and attempt to find which one of these images best matches a region in image #1. Since there is a strong relationship between scale and resolution, one may assume that the scale of the best match corresponds to the resolution ratio between images #1 and #2.

At each scale $s_i$ one-to-one correspondences are determined by direct comparison of the descriptors associated with the interest points. In practice there are 7 such descriptors limited to third-order derivatives of the intensity signal. These descriptors are invariant to image rotation as well as local affine changes of illumination. Two such 7-vector descriptors are compared using the Mahalanobis distance. This distance requires the covariance matrix $\Lambda$ associated with each descriptor. This $7\times7$ matrix encapsulates signal noise, variations in photometry, inaccuracy of interest point location, and so forth. $\Lambda$ is estimated statistically over a large set of image samples. In order to solve as many ambiguities as possible, each one-to-one assignment thus established is checked for local coherence. Namely, for each one of the two points in an assignment we consider their neighbors and check whether the two groups of points in the two neighborhoods are mutually compatible. This local compatibility check based on local geometric distribution has a cost [7] but it is worth the effort because it allows to eliminate spurious matches.

The point matching process just described is applied at 8 scales. Next, we have to evaluate the quality of image-to-image matchings based on these point matches in order to select the scale associated with the best match. We therefore estimate a mapping between the two images as defined by eq. (1) and use robust statistics [2, 14].

Once an approximate scale has been selected using the strategy just described, a robust estimator takes as input the potential one-to-one point assignments, computes the best homography between the two images, and splits the point assignments into two sets: (1) inliers, i.e. points lying in the small region corresponding to the homography mapping of the high resolution image onto the low resolution one and (2) outliers, i.e. points that are either outside this region or mismatched points inside the region.

Commonly used robust estimators include M-estimators, least-median-squares (LMedS), and RANdom SAmple Consensus (RANSAC). In our case, the number of outliers may be quite large. This occurs in particular when the two images have very different resolutions and hence only 20% or less of the low-resolution image corresponds to the high resolution one. Therefore, we ruled out M-estimators because they tolerate only a few outliers. Among the two remaining techniques, we preferred RANSAC because it allows the user to define in advance the number of potential outliers through the selection of a threshold. Hence, this threshold can be chosen as a function of the scale factor. Details concerning threshold selection can be found in [1].

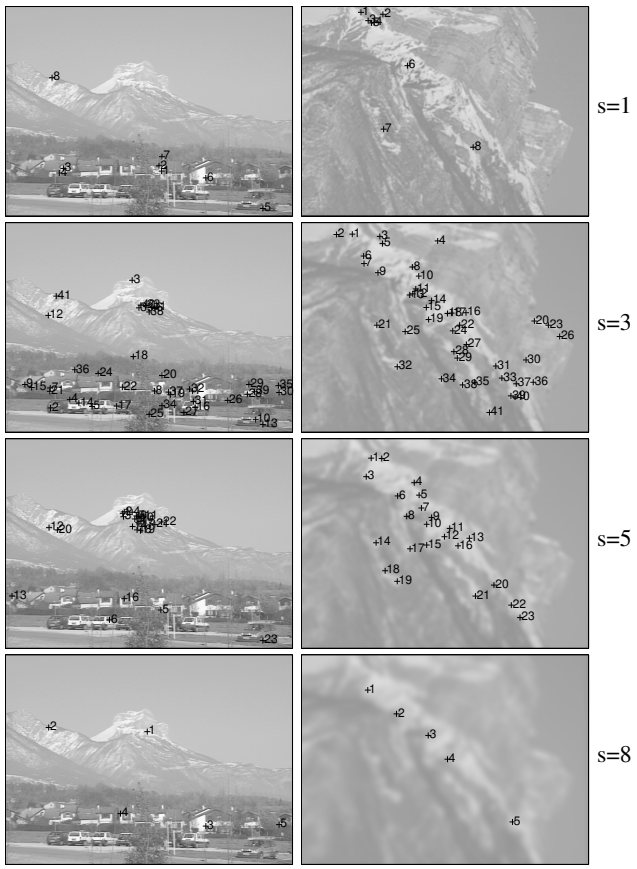


**Figure 3. Point-to-point assignments obtained at four scales.**

## 5 Experiments

The matching strategy just described was applied and tested over a large number of image pairs where the resolution factor between the two images varied from 2 to 6. Here we present three examples. The final result of applying the matching to the pair of Figure 1 is shown in Figure 4.

Let us explain in detail how this type of result is obtained for another example, e.g. Table 1 and Figures 3 and 5. Interest points are first extracted from the low-resolution image at one scale ($s = 1$) and from the high-resolution image at 8 different scales (1 to 8). Therefore, eight image matchings are performed. The result of point-to-point matching is shown on Figure 3 at four different scales: 1, 3, 5, and 8. Obviously scale 3 and scale 5 have the best matches associated with them and scale 5 is a better candidate. Therefore,

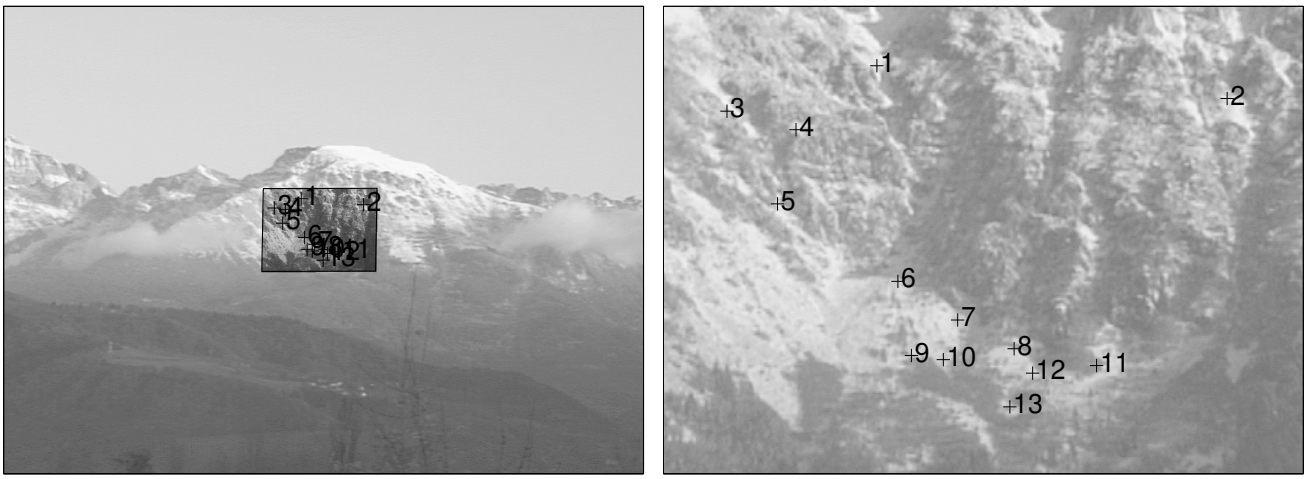


**Figure 4. The high-resolution image is mapped onto the low-resolution one using the homography consistent with 13 point-to-point assignments.**

| Resolution factor | | No. of points | No. of matches | | |
|---|---|---|---|---|---|
| Predicted | Computed | | Initial guess | Inliers | Outliers (%) |
| 1 | 1.3 | 329 | 8 | - | - |
| 2 | 0.7 | 126 | 64 | 4 | 94 % |
| 3 | 1.8 | 64 | 41 | 4 | 90 % |
| 4 | 5 | 31 | 26 | 10 | 62 % |
| 5 | 5 | 25 | 23 | 16 | 30 % |
| 6 | 5 | 18 | 17 | 12 | 29 % |
| 7 | 1.1 | 14 | 14 | - | - |
| 8 | 0.4 | 5 | 5 | - | - |

**Table 1. This table shows, at each scale, the computed resolution factor, the number of points in the high-resolution image, the number of potential matches, the final number of matches, and the percentage of outliers. Notice that scales 5 and 6 yield very similar results.**

it would have been sufficient to run the robust matching algorithm at scale 5 only. In practice we run the latter algorithm at all the scales and displayed the results in Table 1. Thus we can verify that the best match is, indeed, obtained at $s = 5$. Out of 25 points detected at this scale, 23 of them have a potential assignment in the low-resolution image and 16 of them are finally selected by the robust matching technique. The latter rejected 30% of the matches. Finally the homography thus obtained was applied to the high resolution image and this image is reproduced on top of the low-resolution one (cf. Figure 5). A third example is displayed on Figure 6.

## 6 Conclusions

In this paper we presented a new method for matching two images with two different resolutions. We showed that it is enough to represent the high-resolution image in scale-space and we described a one-to-many robust image matching strategy. Key to the success of this method is the scale-space representation of interest points.

In spite of a huge number of publications in the image-matching domain, it seems to us that none of the existing methods is able to deal with large changes in resolution. Here we have been able to match images with a resolution factor of 6. In practice the images shown in this paper were gathered by varying the focal length using the zoom-lens of a digital camcorder. The advent of digital photography opens new fields of applications and we believe that our matching technique will allow the simultaneous exploitation of multiple viewpoints and variable resolution.

## References

[1] G. Csurka, D. Demirdjian, and R. Horaud. Finding the collineation between two projective reconstructions. *CVIU*, 75(3):260–268, 1999.

[2] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6):381 – 395, 1981.

[3] N. Georgis, M. Petrou, and J. Kittler. On the correspondence problem for wide angular separation of non-coplanar points. *IVC*, 16:35–41, 1998.

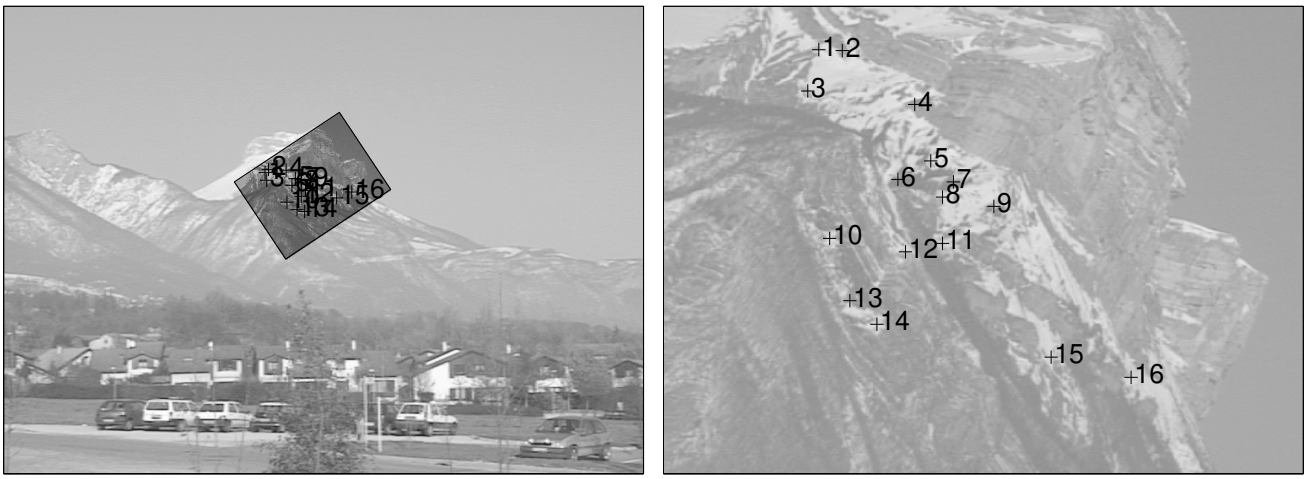


**Figure 5. The final result obtained for the second image pair at scale $s = 5$. All of the 16 matches are correct.**


[4] F. Glazer, G. Reynolds, and P. Anandan. Scene matching by hierarchical correlation. In *CVPR*, 1983.

[5] B. B. Hansen and B. S. Morse. Multiscale image registration using scale trace correlation. In *CVPR*, 1999.

[6] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, 1988.

[7] R. Horaud and T. Skordas. Stereo matching through feature grouping and maximal cliques. *PAMI*,11(11):1168–1180, 1989.

[8] J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–396, 1984.

[9] J. Koenderink and A. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[10] M. S. Lew and T. S. Huang. Optimal multi-scale matching. In *CVPR*, 1999.

[11] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994.

[12] T. Lindeberg. Feature detection with automatic scale selection. *IJCV*, 30(2):79–116, 1998.

[13] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999.

[14] P. Meer, D. Mintz, A. Rosenfeld, and D. Kim. Robust regression methods for computer vision: a review. *IJCV*, 6(1):59–70, 1991.

[15] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *ICCV*, 1998.

[16] L. H. Quam. Hierarchical warp stereo. In *Reading in computer Vision*, pages 80–86. Morgan Kaufman, 1987.

[17] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *PAMI*, 19(5):530–534, 1997.

[18] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *ICCV*, 1998.

[19] T. Tuytelaars, L. V. Gool, L. D'haene, and R. Koch. Matching of affinely invariant regions for visual servoing. In *ICRA*, 1999.


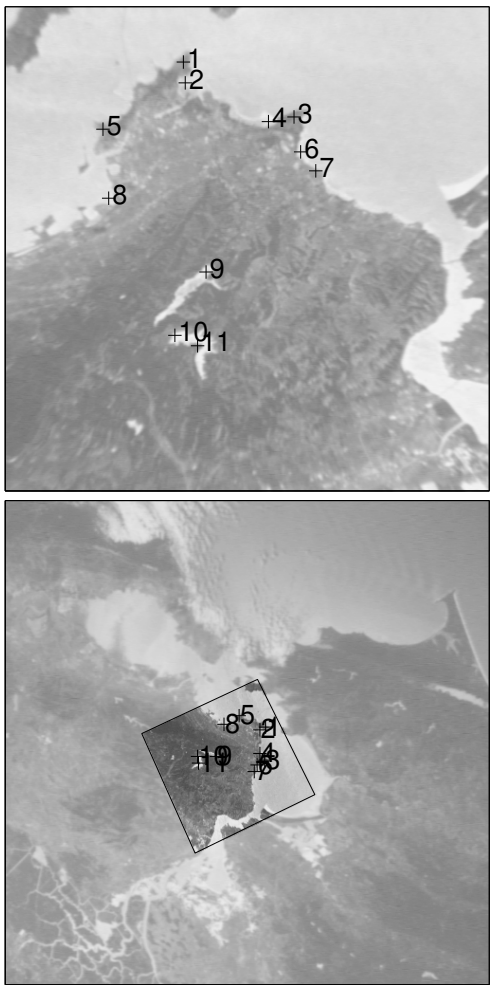


**Figure 6. An other example with $s = 4$**