Object Recognition and Segmentation by Non-Rigid Quasi-Dense Matching

Juho Kannala, Esa Rahtu, Sami S. Brandt and Janne Heikkilä
Machine Vision Group, University of Oulu, Finland
{jkannala,erahtu,sbrandt,jth}@ee.oulu.fi

Abstract

In this paper, we present a non-rigid quasi-dense matching method and its application to object recognition and segmentation. The matching method is based on the match propagation algorithm which is here extended by using local image gradients for adapting the propagation to smooth non-rigid deformations of the imaged surfaces. The adaptation is based entirely on the local properties of the images and the method can be hence used in non-rigid image registration where global geometric constraints are not available. Our approach for object recognition and segmentation is directly built on the quasi-dense matching. The quasi-dense pixel matches between the model and test images are grouped into geometrically consistent groups using a method which utilizes the local affine transformation estimates obtained during the propagation. The number and quality of geometrically consistent matches is used as a recognition criterion and the location of the matching pixels directly provides the segmentation. The experiments demonstrate that our approach is able to deal with extensive background clutter, partial occlusion, large scale and viewpoint changes, and notable geometric deformations.

1. Introduction

This article addresses the problem of recognizing objects in photographs. Object recognition is a wide subject and it can be divided into model-based and appearance-based approaches. Here we consider appearance-based approaches which do not require a specific model of the object. It is assumed that some example images of the object are sufficient for recognition. In addition, we concentrate on recognizing the given object instances in photographs taken under challenging viewing conditions where, for example, the amount of background clutter is large.

Object recognition in the presence of background clutter, occlusion and changing illumination or viewpoint is a difficult problem. Furthermore, the possible deformation of the object between the model and test images provides additional challenge. However, despite the diversity of the problem, recent research has produced many successful recognition approaches [13, 10, 6]. Typically these approaches are local, i.e., they are based on some local viewpoint invariant image features which are matched between the model and test images. The basic building block in the local methods is a region detector which is invariant under viewpoint changes. Several such region detectors have been proposed in the literature [8]. The detectors adapt to the local shape of the intensity surface and hence are able to extract corresponding regions from the model and test images despite the change in viewpoint. Given the detected regions in the model and test images, the most straightforward approach for recognition is to represent the regions with features which allow reliable matching and then use the number of matched features as a recognition criterion [10].

The advantage of the local recognition methods is that they are more tolerant to clutter and partial occlusion than the global approaches [9]. However, even the performance of the local approaches is limited in the presence of extensive background clutter. This is due to the fact that the background produces many incorrect feature matches which disturb the recognition process. In addition, occlusion and large scale or viewpoint changes reduce the probability that a model feature is correctly extracted from the test image. Hence, the combined effect is that the number of matching features is not a reliable recognition criterion since most of the matches are caused by the background. In order to counter these problems a multi-step match-growing strategy has been proposed [2]. This approach consists of alternating expansion and contraction phases which gradually increase the ratio of correct matches. In the expansion phase the current set of region matches is used to construct more matching regions in the surrounding image areas and in the contraction phase some of the mismatches are removed using either a global or local filter. Usually the correctly matched regions grow better than the false ones and this increases the performance of the recognition system [2].

The problem of recognizing a particular object instance in a photograph is closely related to the image registration problem. In fact, the approach in [2] basically searches for the best registration between the model and test images.

However, in some sense the object matching problem described above is more difficult than a traditional non-rigid registration problem where the common area in the images is usually approximately known in advance [1]. Hence, due to the different context, the approaches for solving the registration problem may be quite different [1, 2]. Nevertheless, the match propagation principle used in quasi-dense matching [4, 3] is somewhat similar to the match expansion in [2]. Yet, the quasi-dense approach originates from the context of image matching and, until now, it has not been used for object recognition.

In this paper we propose a new object recognition method which is based on quasi-dense matching. The match propagation technique [4, 3] is here extended to deal with non-rigid image deformations and combined with a new match grouping technique so that it can be applied for object recognition and segmentation. The closest works to that which we report here are [3] and [2]. We use a similar wide baseline match propagation strategy as in [3]. However, the adaptive propagation method proposed in [3] requires that the epipolar geometry between the images is known while the approach proposed here does not have such a limitation. This increases the applicability of the method. Our approach for object recognition and segmentation is conceptually similar to [2]. Yet, the proposed method is more straighforward than that in [2] since there are no repeated contraction phases. In addition, our approach does not use any global constraints and handles the images symmetrically. Hence, it can be applied also in cases where *both* the model image and the test image contain background clutter.

In addition to [2] and [3] the basic idea of growing matches has been used in several works. A region-growing algorithm was proposed in [12] and also the more recent papers [7] and [14] contain somewhat similar ideas. However, these earlier approaches do not address the generic object recognition problem. For example, [7] assumes known epipolar geometry and [14] does not discuss the grouping of matching regions.

2. Background

The match propagation algorithm for quasi-dense matching was proposed in [4] and extended to the wide baseline case in [3]. Since our method utilizes the approach of [3] we briefly review it here.

The basic idea in quasi-dense matching is to compute a large number of point correspondences between two images by starting from a sparse set of initial matches. Affine covariant regions [8] can be used as such initial matches [3]. Hence, the output of the initial matching phase is a set of corresponding points $\{(\mathbf{x}_i, \mathbf{x}'_i)\}_i$ (the centroids of the matched regions) accompanied with the local affine transformation matrices $\mathbf{A}_i$. The initial matches are used as seed points for the match propagation which searches new matches from the surrounding image areas by using the zero-mean normalized cross-correlation (ZNCC) as a similarity measure. The obtained matches are stored in a disparity map which is filled in by iterating the following steps:

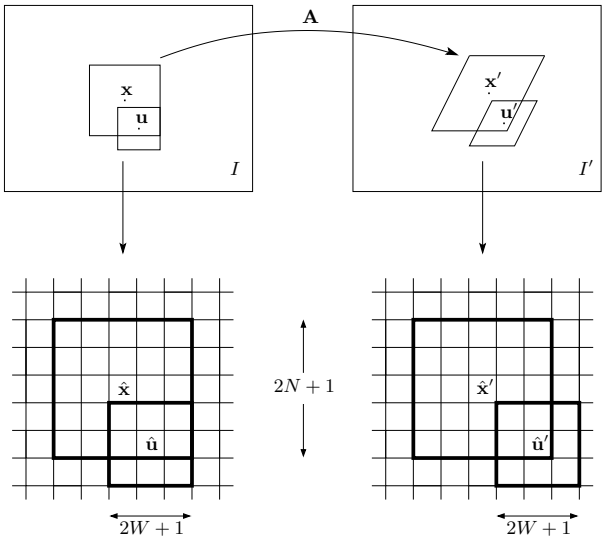


Figure 1. The geometric normalization of local image neighborhoods for a seed match $(\mathbf{x}, \mathbf{x}')$. The pixels $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ in the normalized coordinate frame correspond to the points $\mathbf{x}$ and $\mathbf{x}'$ in the original images. The large black framed windows indicate the search region for new candidate matches. The ZNCC score for the candidate match $(\hat{\mathbf{u}}, \hat{\mathbf{u}}')$ is computed from the smaller black framed windows whose size is $(2W+1)\times(2W+1)$.

(i) the seed point $(\mathbf{x}_i, \mathbf{x}'_i)$ with the highest ZNCC score is removed from the list of seed points

(ii) new candidate matches are searched from the surroundings of $(\mathbf{x}_i, \mathbf{x}'_i)$ by using $\mathbf{A}_i$ for the geometric normalization of local image neighborhoods

(iii) the candidate matches with a sufficiently high ZNCC score are stored in the disparity map and added to the list of seed points

In this manner, the number of correspondences in the disparity map increases until the list of seeds becomes empty.

The geometric normalization process of step (ii) is illustrated in Fig. 1. There the current seed is $(\mathbf{x}, \mathbf{x}')$ and the corresponding affine transformation matrix is $\mathbf{A}$. The local image neighborhoods of $\mathbf{x}$ and $\mathbf{x}'$ are normalized into patches of size $(2N+1)\times(2N+1)$ where from the candidate matches are searched for. The normalization is performed so that the size of the normalized region is $(2N+1)\times(2N+1)$ pixels in the image which locally has a lower resolution. The size of the corresponding region in the other image is determined by the local magnification factor which is either $|\det \mathbf{A}|$ or $|\det \mathbf{A}|^{-1}$. The normalized image neighborhoods are illustrated by the large black framed windows in Fig. 1 and denoted by $\mathcal{N}(\hat{\mathbf{x}})$ and $\mathcal{N}(\hat{\mathbf{x}}')$ in the following.

Given the normalized neighborhoods for the current seed match, the possible candidate matches are given by

$$\mathcal{N}(\hat{\mathbf{x}}, \hat{\mathbf{x}}') = \{(\hat{\mathbf{u}}, \hat{\mathbf{u}}') \mid \hat{\mathbf{u}} \in \mathcal{N}(\hat{\mathbf{x}}), \hat{\mathbf{u}}' \in \mathcal{N}(\hat{\mathbf{x}}'), \\ ||(\hat{\mathbf{u}}' - \hat{\mathbf{x}}') - (\hat{\mathbf{u}} - \hat{\mathbf{x}})||_\infty \leq \epsilon\},$$

where ϵ is the disparity gradient limit [4]. Here we used the value $\epsilon = 1$ which implies that the vectors from the seed point to the candidate point must have approximately the same direction in both normalized coordinate frames. In addition, a candidate match is considered valid only if it is not yet in the disparity map, i.e., neither the pixel closest to $\mathbf{u}$ in image I nor the pixel closest to $\mathbf{u}'$ in image I' is labeled as matched. The ZNCC score is computed for the valid candidate matches using windows of size $(2W+1) \times (2W+1)$ in the normalized domain. Those candidates which exceed a predefined ZNCC threshold z are stored in the disparity map and added to the list of seed points. In the basic propagation mode the new seeds inherit the affine transformation matrix from the current seed. Hence, a seed match is always associated with a local affine transformation which provides the basis for the geometric normalization at each iteration.

Furthermore, in order to prevent mismatching in low-textured regions, a threshold τ may be introduced for the intensity variance of the correlation windows. That is, a seed match is rejected if the intensity variance in its neighborhood is below τ. This is motivated by the fact that the threshold z alone may not be a reliable matching criterion in uniform image areas [4].

In addition to the basic propagation mode described above, an adaptive propagation approach was proposed in [3]. There the idea is to update the estimate of the local affine transformation during the propagation. The adaptation is based on the second order intensity moments and the epipolar geometry. The adaptive propagation mode allows a single seed match to propagate into regions where the local transformation between the images differs from the initial one. However, the adaptation requires that the scene is rigid and the epipolar geometry is known.

As observed in [4], the match propagation algorithm has some desirable properties for image matching. Firstly, the algorithm can be implemented efficiently by using a heap data structure for the fast selection and addition of seed points. Secondly, the algorithm is relatively robust to false matches among the initial seeds. This is due to the best-first propagation strategy which stops the growing of bad seeds in an early stage.

3. Non-rigid quasi-dense matching

In this section we propose an extension to the match propagation technique which allows the propagation to adapt to smooth non-rigid deformations of the imaged surfaces. In addition, we suggest a fast propagation strategy for such cases where a disparity map with a reduced number of point correspondences is desired. Finally, the proposed techniques are illustrated with real image registration examples.

3.1. Non-rigid adaptation

Our non-rigid match propagation method uses the local image gradients and the second order intensity moments to update the estimate of the local affine transformation during the propagation. Hence, unlike in [3], we do not use the epipolar geometry or any other global constraint in matching. Thus, our approach can be used also in cases where the epipolar geometry is not known or the scene is deforming. The details of the method are as follows.

The windowed second moment matrix of the image intensity function f is defined by

$$\mathbf{S}_{f,g}(\mathbf{u}) = \int \mathbf{v}\mathbf{v}^\top f(\mathbf{v}) g(\mathbf{u} - \mathbf{v}) d\mathbf{v}, \tag{1}$$

where the function g is a positive window function. We assume that the intensity function f' and the window function g' are affine transformed versions of f and g so that $f'(\mathbf{u}) = f(\mathbf{A}^{-1}\mathbf{u})$ and $g'(\mathbf{u}) = g(\mathbf{A}^{-1}\mathbf{u})/|\det \mathbf{A}|$. Thus, the coordinate systems in both images are centred to the points under consideration which causes the translational part of the affine transformation to vanish. A change of variables in (1) gives the following transformation rule

$$\mathbf{S}_{f',g'}(\mathbf{u}) = \mathbf{A}\mathbf{S}_{f,g}(\mathbf{A}^{-1}\mathbf{u})\mathbf{A}^\top. \tag{2}$$

Since the origin is the point of interest here, we use the simplifying notations $\mathbf{S}' = \mathbf{S}_{f',g'}(\mathbf{0})$ and $\mathbf{S} = \mathbf{S}_{f,g}(\mathbf{0})$. The positive definiteness of (1) together with (2) implies that

$$\mathbf{A} = \mathbf{S}'^{1/2}\mathbf{R}\mathbf{S}^{-1/2}, \tag{3}$$

where $\mathbf{R}$ is an arbitrary orthogonal matrix. Hence, given $\mathbf{S}$ and $\mathbf{S}'$, the matrix $\mathbf{A}$ can be determined up to a rotation. The idea in [3] is to use the affine transformation of the current seed match to compute the local windows for a new candidate match, and estimate $\mathbf{S}$ and $\mathbf{S}'$ using these windows. Then the affine transformation for the new match is computed by (3) where the remaining rotational degree of freedom is determined from the epipolar lines of the matching points. However, here we determine the rotation by using orientation histograms of local image gradients [6]. That is, we use the histograms to compute the dominant directions of image gradients in the local neighborhoods of the new match. Thereafter, given $\mathbf{S}$ and $\mathbf{S}'$ and a pair of corresponding directions, $\mathbf{d}$ and $\mathbf{d}'$, the affine transformation can be completely determined.

In practice, the computations are carried out in the normalized coordinate frames which are illustrated in Fig. 1. Let us consider the case where $(\hat{\mathbf{x}}, \hat{\mathbf{x}}')$ is the current seed

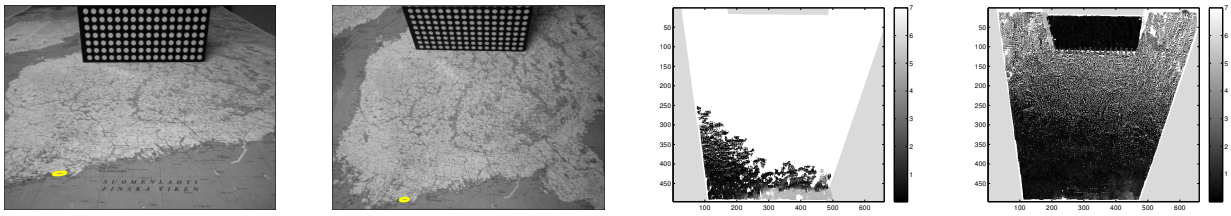


Figure 2. Two images of a rigid scene containing two planes. The matches are shown for both the non-adaptive propagation method (left) and the adaptive propagation method (right) and they are colored according to their Sampson distance from the known homographies [3]. The values over 5 are suppressed to 5, the noncommon image area has grayvalue 6 and the unmatched white area has grayvalue 7. The parameter values used in the propagation were $N=5$, $W=5$, $\epsilon=1$, $z=0.8$, $\tau=0$, $\tau_{\mathrm{a}}=0.25$ and $J=0$.

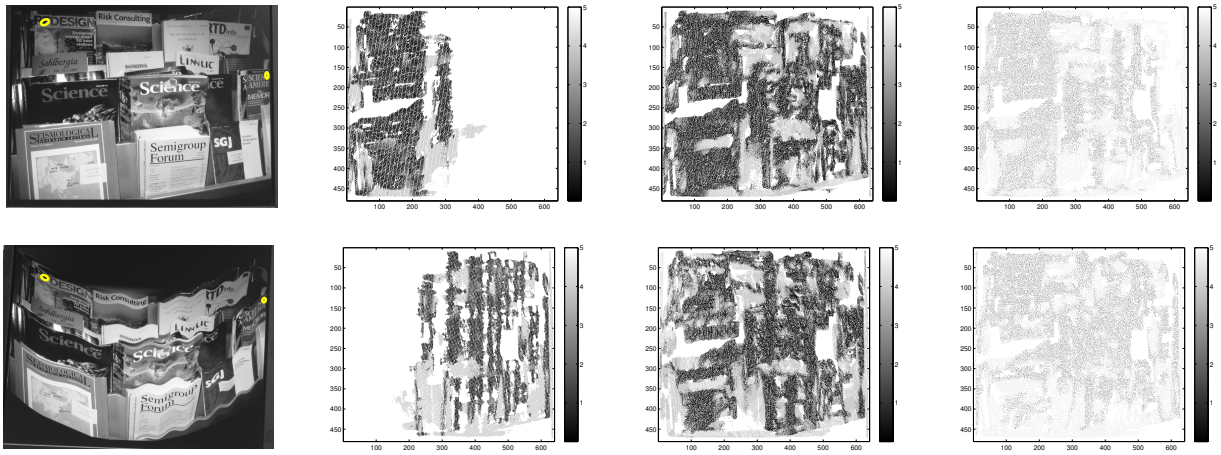


Figure 3. Non-rigid image registration. The distance between the matched point and its true position in the deformed image is used for the color coding. The values over 4 are suppressed to 4. *Top row*: Matches propagated from the seed in the left using the basic mode with $J=0$ (81767 matches, 220 seconds), the adaptive mode with $J=0$ (185027 matches, 627 seconds) and the adaptive mode with $J=1$ (32570 matches, 264 seconds). *Bottom row*: Matches propagated from the seed in the right using the basic mode with $J=0$ (86910 matches, 246 seconds), the adaptive mode with $J=0$ (182726 matches, 622 seconds) and the adaptive mode with $J=1$ (34905 matches, 285 seconds). The matches obtained with $J=0$ and $J=1$ are about equally accurate; the latter just appear more grayish since they are not as dense. The median errors for the matches in the last three columns are 1.2, 1.5 and 1.7 pixels.

and $(\hat{\mathbf{u}}, \hat{\mathbf{u}}')$ is the candidate match under consideration. The aim is to estimate the local affine transformation between the original images at $(\mathbf{u}, \mathbf{u}')$ using the image patches of size $(2W+1)\times(2W+1)$ around $\hat{\mathbf{u}}$ and $\hat{\mathbf{u}}'$. First, the image patches are photometrically normalized and multiplied by a Gaussian window function [3]. Thereafter the moment matrices $\hat{\mathbf{S}}$ and $\hat{\mathbf{S}}'$ are computed from the patches and transformed to the original coordinate frames by $\mathbf{S}=\hat{\mathbf{S}}$ and $\mathbf{S}' = \mathbf{A}\hat{\mathbf{S}}'\mathbf{A}^\top$, where $\mathbf{A}$ is the transformation used in the normalization process. Second, the dominant gradient directions $\hat{\mathbf{d}}$ and $\hat{\mathbf{d}}'$ are computed from the patches and transformed back to the original frames, i.e., $\mathbf{d} = \hat{\mathbf{d}}$ and $\mathbf{d}' = \mathbf{A}\hat{\mathbf{d}}'$. The gradients are computed by convolving the patches with the derivatives of a Gaussian filter and their magnitude-weighted orientations are stored in a histogram with 36 bins. The histogram is smoothed and the dominant gradient direction is found by fitting a parabola to the three values closest to the highest peak in the histogram [6]. Finally, the affine transformation for the new seed match $(\mathbf{u}, \mathbf{u}')$ is computed from $\mathbf{S}, \mathbf{S}'$ and $\mathbf{d}, \mathbf{d}'$ as described above.

The advantage of performing the computations in the normalized frame is that the values of the Gaussian window function and the derivatives of the Gaussian filter can be calculated in advance since the window is always the same. In addition, due to the separability of the isotropic Gaussian filter the gradient can be computed efficiently with two 1D convolutions. In summary, the image interpolation, illustrated in Fig. 1, is done only once at each propagation step and this allows efficient propagation also in the non-rigid case.

Finally, in order to make the adaptation more robust we introduced a threshold τ_{a} for the minimum intensity variance in the local patches. That is, the adaptation is performed only if the variance of the image intensity in the neighborhood of a new match is sufficiently high. Further-

more, if either one of the orientation histograms is very flat so that the peaks can not be reliably identified, the adaptation is not performed.

3.2. Fast propagation by jumping

Due to the high resolution of the images the number of quasi-dense matches may be unnecessarily large for applications. For example, points lying very close to each other do not provide much additional information in surface reconstruction but make the computational cost of subsequent processing high [5]. Hence, a resampling strategy was used in [5] where a reduced set of point correspondences was computed by fitting an affine transformation to several quasi-dense matches inside small image patches. Here we suggest an alternative approach where the match propagation algorithm directly produces a reduced but uniform set of correspondences. This is achieved by modifying the propagation algorithm of Section 2 so that in step (iii) only such seed matches are accepted whos $(2J+1)\times(2J+1)$ neighborhood does not already contain matches. The modification causes the propagation to take larger steps and hence proceed faster. The parameter J determines the jump size. Although a very large value of J may reduce the adaptivity of the method, we found that usually a small value, such as $J=1$, can be safely used for faster propagation.

3.3. Examples

The non-rigid matching method is illustrated with the examples in Figs. 2 and 3. First, in Fig. 2, we have the same image pair which was used in [3] and is available at [15]. There the scene contains two planes for which the homographies between the views are known so that the matches can be evaluated. The match propagation was started from a single seed region which is illustrated by the ellipses in Fig. 2. The last two columns in Fig. 2 show the propagation result obtained by using both the basic propagation mode and the adaptive propagation mode. Only the adaptive method is able to proceed into such image regions where the local transformation differs from the initial one. The obtained result is comparable to [3] although here the epipolar geometry was not used to either constrain the matching or assist the adaptation. This shows that our adaptation method is stable enough when applied to a rigid scene.

In the second experiment, illustrated in Fig. 3, we introduced a non-rigid deformation by displaying an image and its deformed version on a flat screen display and taking a photograph of both images. The homographies between the images and their photographs were determined by displaying a calibration pattern on the screen. Since the artificial image deformation and the homographies were known we calculated the geometric transformation from the first photograph to the second one and used it as a ground truth for evaluating the quasi-dense matches. The results are illustrated in Fig. 3. It can be seen that the non-rigid adaptation clearly improves the matching. Both of the illustrated seed regions propagate well and most of the false matches are located in low-textured image regions where the propagation could be further prevented by increasing the threshold τ. The computation times reported in Fig. 3 were obtained by our current Matlab/MEX implementation which is not optimal for efficiency.

4. Segmentation and recognition

This section proposes an approach which utilizes the quasi-dense matches for segmenting and recognizing common objects in two images. The approach is based on grouping the matches into geometrically consistent components which are supposed to lie on smooth surfaces representing the objects. In addition, we introduce a measure for evaluating the reliability of the obtained segments.

4.1. Match grouping

The match grouping, illustrated in Fig. 4, has two phases. First, a set of tentative segments is made by forming a single group from all the matches which have grown from the same initial seed match. Second, the neighboring groups are merged if the local affine transformations on the boundaries of the groups are sufficiently similar with each other and in good agreement with the spatial arrangement of the groups.

The tentative grouping phase is illustrated in the first column of Fig. 4. On the grounds of the properties of the match propagation, described in Section 3, it is reasonable to assume that the matches originated from a single seed lie on a smooth surface. Here these surfaces are considered as objects and, as can be seen from Fig. 4, the tentative segments are indeed located on single objects. However, there may still be several segments on a particular object and the remaining task is to merge such segments.

The grouping of neighboring segments is based on considering such sets of three nearby matches where all the matches are not yet in the same segment. The studied sets are found by applying the Delaunay triangulation algorithm [11] to all matches in both images and then discarding such triangles where all the vertices already belong to the same group. The geometric consistency of the remaining triangles is examined, as described below, and for each consistent triangle the segments associated to the vertices are merged. The triangles before and after the consistency check are illustrated in Fig. 4.

The geometric consistency of sets of three matches is evaluated as follows: (a) the centroids of the three matching points are translated to the origin in both images, (b) the translated points from the image with a larger local scale

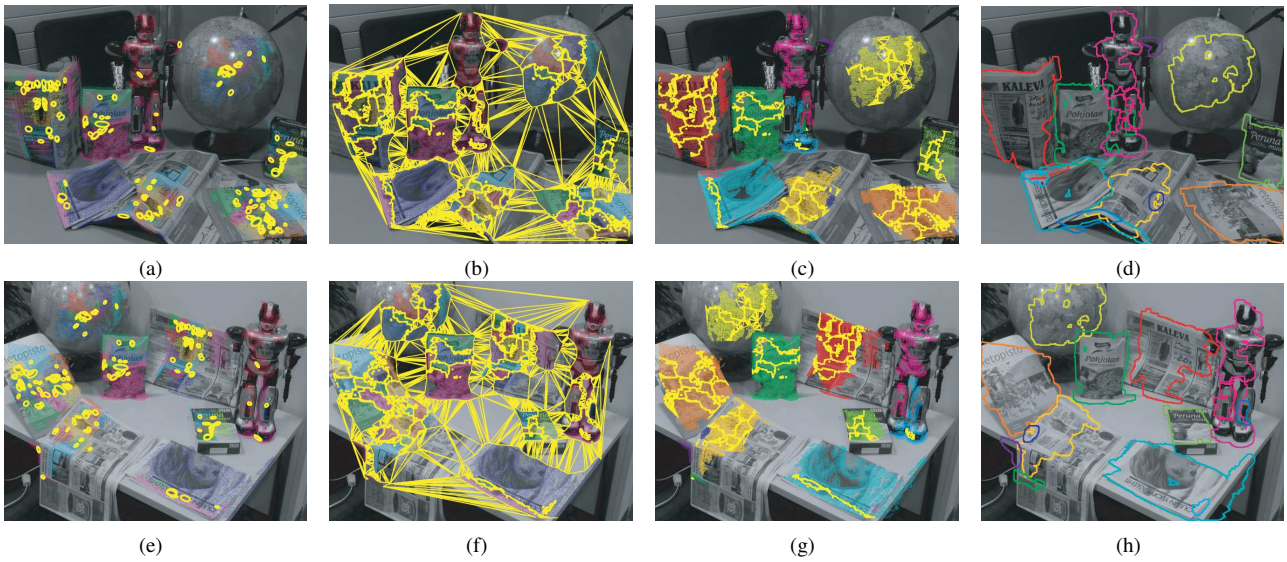


Figure 4. The recognition and segmentation process illustrated step by step. (a) and (e) show the initial seeds (yellow ellipses) and the quasi-dense matches; matches originating from different seed points are plotted with different colors. (b) and (f) show the result of the Delaunay triangulation. (c) and (g) show the surviving triangles after the consistency check. (d) and (h) show the final segmentation results.

are mapped to the other image using the affine transformation matrices associated with each match during the propagation, (c) the Euclidean distances from the resulting nine points to their translated correspondences are computed, (d) the maximum of all these distances (i.e. the maximum displacement) is used as a measure for the geometric consistency of the three matches. The segments joined by at least one set of three matches, whose maximum displacement is below a predetermined threshold, are merged. The threshold for the maximum displacement was 5 pixels in the examples of Figs. 4 and 5.

The Delaunay triangulation is an efficient way for joining neighboring segments. The computational complexity of the algorithm is $O(n \log n)$ where n is the number of points. Furthermore, performing the triangulation in both images allows also to merge segments which are separated by a mismatched segment inbetween them in either one of the images. For example, the mismatch could be caused by an occlusion which is present in the other image. However, it is more unlikely that two segments of the same object would be completely isolated from each other in both images by mismatched segments. Hence, our local approach for segmentation is tenable in practice.

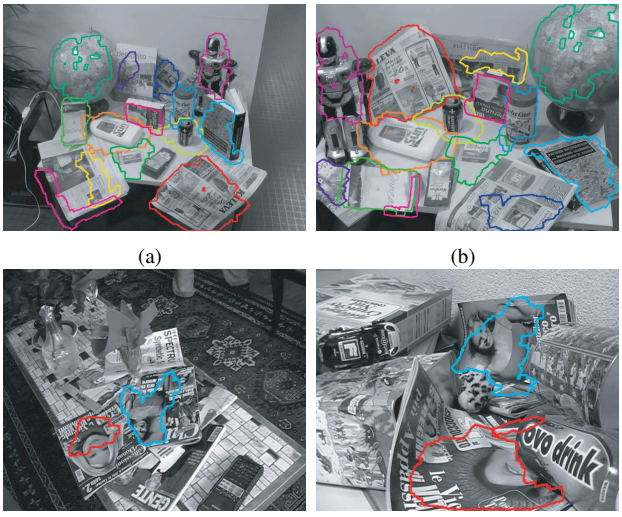
4.2. Recognition

After the match grouping the recognition system has to determine which segments represent real objects and which are false matches. To address this problem we evaluate the obtained segments by computing their correlation weighted areas in both images (i.e. the area covered by the correlation window of each match in the segment is weighted with the cross-correlation score obtained during the match propagation) and taking the minimum of these as a reliability measure. Typically the false matches have smaller coverage which implies that they are not considered as reliable as the correct ones. For example, in the last column of Fig. 4 we have illustrated all the segments obtained at the grouping phase for our example image pair. The eight segments with the largest correlation weighted areas correspond to the eight objects in the images.

Often the recognition task is such that we are given two images of which the first one contains a model object on a uniform background and the other is a test image and we are asked to determine whether the model object is present in the test image [2]. Our approach above is directly applicable also in this kind of recognition task. However, in this case we may directly use the correlation weighted area in the model image as a recognition criterion since we know that the model image contains nothing else than the object. Furthermore, if the model object is segmented from its uniform background, we may multiply the correlation weighted area with the relative coverage of the object area. This makes the models of different sizes more comparable. However, in general the model objects need not to be segmented, in contrast to the approach in [2].

5. Experiments

In this section we present experiments which demonstrate our approach in recognition and segmentation tasks. The first experiment in Fig. 5 illustrates the general case where the task is to find and segment the common objects in two images. In the second experiment, shown in Figs. 6-



(a) (b) (c) (d)

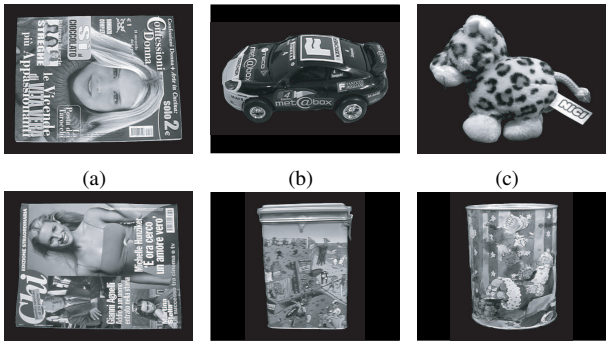
Figure 5. The segmentation results in the case of two images with several common objects and background clutter.

8, the recognition performance is evaluated using the same dataset as in [2].

The first image pair in Fig. 5 contains 14 common objects. The 14 segments, which were considered the most reliable matches, are illustrated with different colors. There is one false match among the segments and it is the L-shaped region in magenta. The only true object missing is the tea bag which was not found due to few seeds on it. One can observe that some obtained segments include parts of the table in the background. This is because the table is the same in both images and it is not completely uniform in intensity. Yet, if necessary, the matching in relatively uniform regions could be further prevented by increasing the thresholds for the propagation. Overall, the segmentation result is fairly accurate despite the occlusions and deformations present.

As the second test pair in Fig. 5 we have two images taken from the ETHZ toys dataset [16]. Despite significant background clutter in the images the method correctly found the magazines as the two most reliable matches. The segmentation of the magazine indicated by the blue line in Fig. 5 is almost perfect, except for the small strongly folded part at the lower left corner. The other segmentation illustrated with the red line is slightly less accurate leaving out only the lower left corner, which is strongly folded and has some illumination distortions.

In the last experiment we performed the same object recognition task as in [2] using the ETHZ toys dataset [16]. In this dataset there are 9 model objects and 23 challenging test scenes where one or more model objects are present. Figs. 6 and 7 illustrate some samples of the model and test images. The recognition task is to determine which model objects are present in the test scenes and to find the corresponding segmentations.



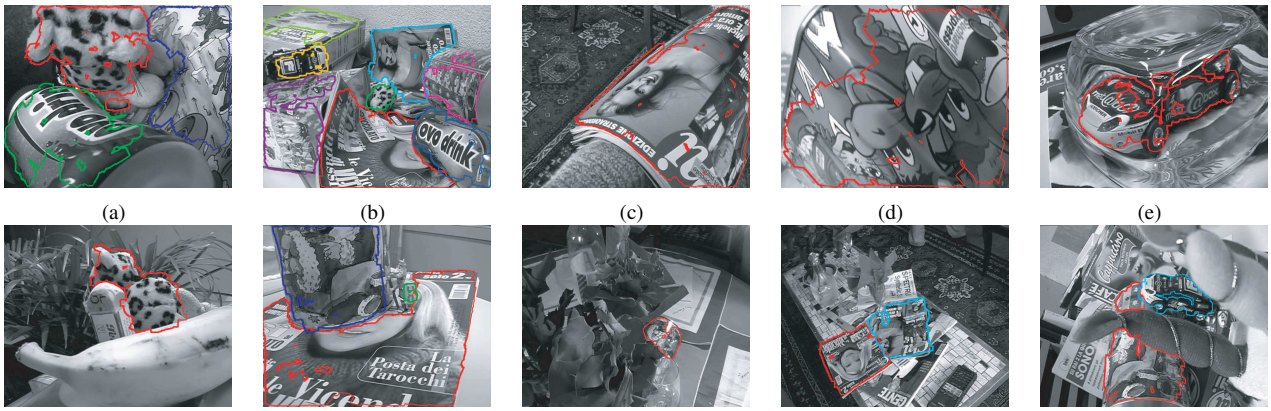
(a) (b) (c) (d) (e) (f)

Figure 6. Six images of model objects in the ETHZ toys dataset.

The recognition was performed by applying our method to all model and test image pairs. That is, the match propagation was first performed with the values $N = 12$, $W = 8$, $\epsilon = 1$, $z = 0.85$, $\tau = 0.45$, $\tau_{\mathrm{a}} = 0.7$ and $J = 1$ whereafter the matches were grouped using a threshold of 20 pixels for the maximum displacement criterion. The reliability of the obtained segments was finally measured by their correlation-and-coverage-weighted area as described in Section 4.2. The weighted area of the most reliable segment was used as an evidence that the model under consideration is present.

Some of the objects in the dataset had more than one model view and in this case we computed the total evidence as a sum of the evidences of the views. In addition, if the magnification factor of the best matching segment was below 0.06 the system was set to give a negative recognition result independently of the evidence value. This removes such false detections where the region in the model image is very heavily downscaled. (The downscaling leads to a loss of details and can thus cause erroneously high evidence values when the object is not present at all.)

In order to quantify the recognition performance we computed the ROC curve by altering the decision threshold for the evidence value. Fig. 8 illustrates the resulting curve (blue) and also the one produced in the same experiment in [2] (black). It can be seen that our method gives almost as good performance as the method in [2] without using any color information. It is likely that the lack of color information explains much of the difference between the curves. We demonstrate the effect of incorporating color by taking it into account in a naive way, i.e., by computing the 10×10 normalized color histogram for the best matching segment in both images and then dividing the original evidence with the intersection of the histograms. The ROC curve with the new evidence is also plotted in Fig. 8 (red). It can be seen that already this simple incorporation of color gives results comparable to [2]. However, it would be relatively straightforward to utilize color directly in the match propagation phase and this might improve the results further.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 7. Examples of segmentation results with the ETHZ toys dataset.

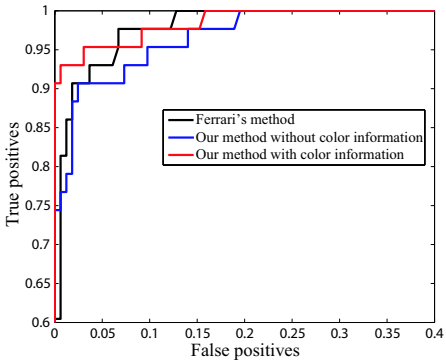


Figure 8. ROC plot for the ETHZ toys dataset.

6. Conclusion

In this paper we proposed a non-rigid match propagation method which adapts to smooth deformations of the imaged surfaces using only local properties of the images. The experimental results show that the presented method can be successfully used for image registration in the presence of notable geometric deformations. In addition, we proposed a new approach for match grouping which directly utilizes the local affine transformation estimates obtained during the match propagation. The grouping allows to use the quasi-dense approach for segmenting the common objects from the images. Furthermore, the groups of quasi-dense matches can be directly used for object recognition. The object recognition results obtained with a commonly available dataset show that our approach is able to deal with challenging viewing conditions, such as occlusion, clutter, geometric deformations and large viewpoint changes.

Aknowledgements

We are grateful to Vittorio Ferrari for providing his data for the experiments. We would also like to thank the anonymous reviewers for their feedback.

References

[1] A. Bartoli and A. Zisserman. Direct estimation of non-rigid registrations. In *BMVC*, 2004.

[2] V. Ferrari, T. Tuytelaars, and L. Van Gool. Simultaneous object recognition and segmentation from single or multiple model views. *IJCV*, 67:159–188, 2006.

[3] J. Kannala and S. S. Brandt. Quasi-dense wide baseline matching using match propagation. In *CVPR*, 2007.

[4] M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. *TPAMI*, 24(8):1140–1146, 2002.

[5] M. Lhuillier and L. Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *TPAMI*, 27(3):418–433, 2005.

[6] D. Lowe. Distinctive image features from scale invariant keypoints. *IJCV*, 60:91–110, 2004.

[7] Z. Megyesi and D. Chetverikov. Enhanced surface reconstruction from wide baseline images. In *3DPVT*, 2004.

[8] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *IJCV*, 65:43–72, 2005.

[9] H. Murase and S. Nayar. Visual learning and recognition of 3D objects from appearance. *IJCV*, 14:5–24, 1995.

[10] S. Obrdzálek and J. Matas. Object recognition using local affine frames on distinguished regions. In *BMVC*, 2002.

[11] J. O'Rourke. *Computational geometry in C*. Cambridge, 2nd edition, 1998.

[12] G. P. Otto and T. K. W. Chau. Region-growing algorithm for matching of terrain images. *Image and Vision Computing*, 7(2):83–94, 1989.

[13] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *TPAMI*, 19(5):530–535, 1997.

[14] A. Vedaldi and S. Soatto. Local features, all grown up. In *CVPR*, 2006.

[15] http://www.ee.oulu.fi/∼jkannala/quasidense.html.

[16] http://www.robots.ox.ac.uk/∼ferrari/datasets.html.