

PLAUSIBLE IMAGE MATCHING: DETERMINING DENSE AND SMOOTH MAPPING BETWEEN IMAGES WITHOUT *A PRIORI* KNOWLEDGE

SHUNTARO YAMAZAKI

Digital Human Research Center

National Institute of Advanced Industrial Science and Technology (AIST)

Water Front 3F, 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan

shun-yamazaki@aist.go.jp

KATSUSHI IKEUCHI

Industrial Institute of Science, University of Tokyo

4-6-1 Komaba, Meguro, Tokyo 153-8505, Japan

ki@cvl.iis.u-tokyo.ac.jp

YOSHIHISA SHINAGAWA

Beckman Institute, University of Illinois

2021 Beckman Institute, University of Illinois

405 N. Mathews, Urbana, Illinois 61801, USA

sinagawa@uiuc.edu

This paper presents a method for automatic determination of dense and smooth mapping between two images without *a priori* knowledge of either the camera pose or the objects in the images. We designed an algorithm to find the mapping between a pair of arbitrary images, and accomplish automatic image morphing. In order to extract image features which look natural to human, we use a set of linear filters similar to those that are used in early vision. Then the derived vector fields consisting of filter responses are matched with each other through a minimization of the cost function which expresses the similarity of transformed images and mapping smoothness, in a multiresolutional hierarchy. Since the cost function in general is highly nonlinear, we avoid excessive distortion in the estimated mapping by providing a local convexity of mapping in nonlinear optimization. In this paper, a variety of experimental results are discussed for various data sets, including images of rotating objects, static objects, human faces and texture patterns, to demonstrate the performance of the proposed method.

Keywords: Unconstrained image matching; plausible image matching; rotated Gaussian derivatives; convexity constraint; nonlinear optimization.

1. Introduction

Image matching is a generic process of finding correspondence between two or more different images. This is a fundamental task in virtually any image analysis in which

the final information is obtained from the combination of different imaging sensors' output. When we want to integrate or compare the images of different objects, taken at different times, from different viewpoints, and/or through different sensors, it is essential that the images are geometrically aligned.

If a scene is common among given images, or at least the objects in the images have semantic correspondence, then we can define physically correct matching between the images. The estimation of this type of matching is referred to as *image registration*.^{5,30} A broad range of techniques of image registration have been developed for various type of data in many application areas, such as, remote sensing, multimodal medical imaging, cartography, virtual world heritage and computer vision. The techniques of image registration have achieved considerable success and are regarded as essential in these fields.

With rapid progress in image processing technologies and computational environment, certain demands for a broader scope of image matching techniques have recently been increased among entertainment communities. A typical application of this type of technique is image morphing³ which is a technique typically used as an animation tool for smoothly and gradually blending one image with another. Smooth image interpolation also brings great benefits to image-based rendering techniques from a sparse image set.

In this paper, we propose a method of image matching that can find correspondences without *a priori* knowledge of either different factors affecting imaging process or objects included in the scene. Our method can determine image correspondence that looks natural to human even if the objects in the captured images are different. Since it is difficult, if not impossible, to define correct correspondence between the images, we determine an optimal correspondence with respect to image features to which human visual system is sensitive. We refer to this technique as *plausible image matching*.

Our proposed method can determine dense correspondence between two images through optimization. When no information on what appears in given images is available, the difficulties in solving the correspondence problem are mainly the following two. First is the selection of the features by which the desirable correspondence between images is defined. Since the corresponding points may differ in their image intensities, it is essential to define the similarities between the images in a plausible manner. Second is on how to prevent the optimization process from being trapped into a local minimum when the images come with numerous features and distant correspondences.

The rest of this paper is organized as follows. After summarizing related work in Sec. 2, the problems to be solved are formulated as the minimization of a certain cost function in Sec. 3.1. The concrete objective function is defined in Sec. 3, and then the method of minimizing the function, preventing the mapping from being distorted is proposed in Sec. 3.4. Details of the algorithm are presented in Sec. 4, and the results obtained by applying our proposed method to various data sets are shown in Sec. 5. Finally the conclusion is drawn in Sec. 6.

2. Related Work

As mentioned in Sec. 1, a broad range of image registration techniques have been developed for various types of data and conditions. According to comprehensive surveys published in 1992 by Brown⁵ and in 2003 by Zitová,³⁰ most methods of determining a match between two images consist of the following three steps: feature extraction, feature matching and transform model estimation. In this section we present brief explanations of these processes and introduce related work.

2.1. Feature extraction

Algorithms for image matching extract the information from within the images that will be used for matching. The image intensity values can be used directly as a clue for solving the correspondence problem, especially in area-based image matching as described in the next section. When the objects in a given image have distinct local structure, matching algorithms can use the image features to achieve robust and effective estimation. The features should be distinct, spread all over the image and efficiently detectable in all given images.

Conventional image features are broadly classified into three types: salient points, lines or regions in the images. Point features can be extracted by detecting line intersections,²³ the most distinctive points with respect to a specified measure¹³ and corners.²⁷ In computer vision applications, corners are commonly used as features in given images. A comprehensive survey of corner detectors was presented by Zheng *et al.*²⁹ Line features can be the representations of general line segments in given images and object contours. Standard edge detection methods, like the Canny detector⁶ or a detector based on the Laplacian of a Gaussian,¹⁷ are employed for line feature detection. Region features can be used to detect the projection of high contrast closed-boundary regions and patterns. A basic approach to region detection is image segmentation.¹⁹ On the other hand, Alhichri and Kamel¹ proposed the idea of virtual circles, which can select region features that are invariant with respect to change of scale.

Recently, some researchers presented more intuitive techniques that apply non-linear operations over image pixels. Smith and Brady²² proposed the SUSAN feature detector based on counting the number of interesting pixels within a close neighborhood. Shinagawa and Kunii proposed the critical point filters (CPF)²¹ that extract local maximum and minimum of pixel intensities, repeatedly in a multi-resolutional hierarchy.

In many applications, such as remote sensing and computer vision, the images usually contain enough distinctive and easily detectable features mentioned above. However, it is difficult to justify the choice of the features used for matching images when the objects in the images are unknown. In this paper, we propose a technique for tackling this difficulty, taking into account the process of image recognition in the human visual system.

2.2. Feature matching

The extracted image features are matched so that the similarity between corresponding features is maximized. When the features are not so descriptive that the mapping cannot be determined uniquely using only the features, additional constraints, such as the regularization of a mapping function, are also taken into account.

When the local structural information in given images is more significant than the information carried by the image intensities, feature-based matching methods are applied. A comprehensive survey is presented by Zitová.³⁰ These methods allow the alignment of images of a completely different nature, and can handle complex inter-image distortions. It is, however, often the case that the respective features are hard to detect and therefore unreliable. It is also difficult to have discriminative and robust feature descriptors that are invariant to all assumed differences between the images.

On the other hand, area-based matching techniques can be used to deal with the images without attempting to detect distinctive features. Windows of predefined size or even entire images are used for correspondence estimation that is based on correlation maximization,²⁰ phase correlation maximization,⁴ template matching⁷ or mutual information maximization.²⁵ Area-based matching methods can be accelerated by pyramidal image representations²⁴ to find the maximum of a similarity measure.

Our matching algorithm does not impose any assumption on objects in given images, and is based on area-based methods. However, the similarity between corresponding pixels can be calculated without an explicit window operation, since information on pixel intensity distribution around a pixel is represented as a feature vector before the images are matched. In addition, the estimation of image matching is carried out in a multiresolutional hierarchy.

2.3. Matching estimation

Once a strategy of image matching is decided, image matching algorithms try to find the mapping function that transforms one of the images to overlay it over the other one through the maximization of a certain similarity metric. The task to be solved consists of choosing the type of mapping function and its parameter estimation.

When the transformation between given images can be represented by a global polynomial mapping of low degree, the mapping is unambiguously determined from a small set of corresponding points, through least square fitting. Goshtasby *et al.*⁸ proposed the weighted least square method, and improved matching accuracy for images which have local geometric distortion. Locally varying geometric distortion can be handled by radial basis functions, such as, multiquadric functions,¹⁶ Wendland's function¹⁸ or thin-plate splines.²⁶ These methods are suitable for

estimating parameters of such simple transformations as similarity transformation, affine transformation and projective transformation.

A more general, complex matching has to be modeled using nonrigid transformation. Elastic matching² regards the images as pieces of a rubber sheet, and models the estimation of geometric deformation as the search for the best parameter of the sheets. Certain external forces stretching the image and internal forces defined by stiffness or smoothness constraints are applied on the sheets to bring them into alignment with the minimal amount of bending and stretching. Another approach to the registration of images with considerable complex distortions is the optical flow method¹¹ which can estimate relative motion between images using only image intensity as a constraint. Shinagawa and Kunii proposed a more intuitive method for automatic determination of image correspondence.²¹ They formulated the correspondence problem as an estimation of grid deformation, and solved the problem by local exhaustive search in a multiresolutional hierarchy. Habuka and Shinagawa⁹ improved the algorithm taking into account the consistency between forward and backward mappings.

We propose an algorithm for matching estimation based on the framework proposed by Shinagawa and Kunii. Our algorithm is improved for the following two points. First, ours adopts an iterative optimization approach aiming for better matching estimation. Second, we avoid excessive distortion in the estimated mapping by providing a local convexity of mapping in the nonlinear optimization.

3. Plausible Image Matching

In this section, the method of determining dense correspondence between only two images is proposed. As mentioned in Sec. 1, the keys to solving this problem are: how to select features by which the desirable correspondence between images is defined and how to prevent the mapping from being converged into a local minimum without defects such as strong distortion when the images come with numerous features and distant correspondences.

3.1. Problem formulation

Let $I : \mathbb{D} \rightarrow \mathbb{V}$ be an image color distribution, where \mathbb{D} is a set of pixel coordinates and \mathbb{V} is the space of pixel values. The problem to be solved is then formulated as the estimate of optimal mapping $m : \mathbb{D} \rightarrow \mathbb{D}$ from images I_{src} to I_{dst} in the sense that an appropriate measure $\varepsilon_{\mathbb{D}}^2$ which expresses the difference between $I_{\text{src}}(m(\cdot))$ and $I_{\text{dst}}(\cdot)$ is minimized. We allow the images to be upsampled, that is, the resolution of \mathbb{D} may be higher than that of pixel grids. When the resolutions of image pixels and mapping between them do not match, we interpolate pixel values by bilinear interpolation. In addition, mapping is restricted to a bijection, that is, m^{-1} is uniquely defined at any point in \mathbb{D} .

Hereafter, mainly the method of matching two-dimensional gray images is discussed for simplicity.

3.2. Feature extraction

In order to solve correspondence problems, the algorithm attempts to match features in one image with corresponding features in the other. The intensity of pixels in an image should not be used directly since the corresponding pixel may have different intensities in the concerned problem. Unlike solving other correspondence problems in computer vision like stereo photogrammetry, we may have to find the correspondence between points on two unrelated objects. Nevertheless, matching is expected to appear natural to humans. This observation implies that it is helpful to apply the feature detector actually used in human vision.

Similar to the related work in the field of pattern recognition,¹⁵ we use partial derivative of a Gaussian which well approximates visual receptive fields in early vision.²⁸ When the two-dimensional images are matched, we use a set of rotated Gaussian derivative filters $K_{\sigma,n,\theta}$ defined as

$$K_{\sigma,n,\theta}(x, y) = G_{\sigma}^{(n)}(x \cos \theta - y \sin \theta) \times G_{\sigma}^{(0)}(x \cos \theta + y \sin \theta) \quad (1)$$

where $G_{\sigma}^{(n)}$ is the n th derivative of a Gaussian with standard deviation σ .

If some of the filters in a filter bank have linear dependency, filter response may be redundant. By applying singular value decomposition to the matrix consisting of the vectors of filter kernels, the number of bases required for n th Gaussian derivative will be $n + 1$.¹² Based on this, we used 24 filters consisting of one Gaussian, two first derivatives and three second derivatives of Gaussian with four different spatial scales, as shown in Fig. 1.

At the beginning of the image matching, a filter bank $F = \{F_i\} (i = 1, \dots, N_f)$ composed of $K_{\sigma,n,\theta}$ with N_f different parameter settings is applied to each of the given images, and then the filter responses are stored in the form of vector fields with the same size as the original images. We call these vector fields *feature vector field*. The correspondences between images are determined only by the fields.

3.3. Feature matching

Optimal mapping m from I_{src} to I_{dst} is estimated by maximizing the similarities between $I_{\text{src}}(m(\cdot))$ and $I_{\text{dst}}(\cdot)$ under an appropriate constraint on the regularity of the mapping. This problem can be formulated by the objective function concerning the similarity of transformed images and mapping smoothness.

The first criterion refers to the matching quality of the entire region of the images. That is, the objective function ε_D^2 for the similarity is defined as

$$\varepsilon_D^2 = \int_{p \in \mathbb{D}} \sum_{i=1}^{N_f} \rho_D((F_i * I_{\text{src}})(p) - (F_i * I_{\text{dst}})(m(p))) \quad (2)$$

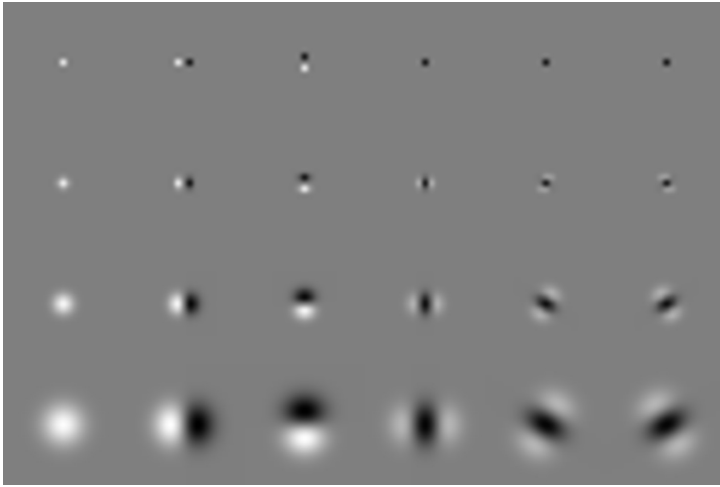


Fig. 1. The filter bank used to generate feature vectors. We used 24 filters composed of the original Gaussian, and first and second partial derivatives of Gaussian, with four scales. The filters have zero mean and each of them is divided by its L_1 norm for scale invariance.

where the operator $*$ denotes the convolution and ρ_D is the M-estimator¹⁰ which is robust for the perturbations of feature vectors. In our implementation, Lorentzian function

$$\rho(x) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) \tag{3}$$

is used as the M-estimator. ρ_D is defined by the sum of Eq. (3) with the scale $\sigma = \sigma_D$ for all elements of $F_i * I$.

The second criterion is based on the regularity of the mapping. Although the feature vectors defined in Sec. 3.2 may have sufficient information for determining the correspondence uniquely, regularization is essential since it is impossible to determine the mapping where the image intensity is constant. The objective function ε_S^2 reflecting the smoothness of the mapping $m = (m_1, \dots, m_{\dim\mathbb{D}})$ can be defined as

$$\varepsilon_S^2 = \int_{p \in \mathbb{D}} \sum_{d=1}^{\dim\mathbb{D}} \rho_S(\nabla m_d(p)) \tag{4}$$

where ρ_S is the robust M-estimator that returns the sum of Lorentzian with scale σ_S for all elements of ∇m_d in our implementation.

By combining Eqs. (2) and (4) with the weighting parameter α^2 , the objective function ε^2 to be minimized is defined as follows.

$$\varepsilon^2 = \varepsilon_D^2 + \alpha^2 \varepsilon_S^2. \tag{5}$$

The parameter α^2 is used to regularize mapping in the region where the color is nearly constant; hence, α^2 can be small. The parameter σ_S should be sufficiently

large since the difference of the features is the only cue to determining the correspondence in our problem. The parameter σ_D is heuristically selected in our experiments, for instance, 5% of image width.

3.4. Matching estimation

In general, it is difficult to find the optimal parameter α^2 since the difference between images can be so large that global matching may cause local distortions such as twists and flips in the mapping which violate our assumption on bijectivity. We prevent the mapping from being distorted by imposing restrictions for the mapping to preserve its local convexity.

When the four vertices of a small square in I_{src} are mapped into I_{dst} by m , the possible configurations of these four vertices are as shown in Fig. 2 since different points are mapped to different points due to the bijectivity of the mapping. In the determination of mapping m , we ignore patterns (c) and (d) because bijectivity is broken either inside or near the quadrilateral. Note that, even if we ignore the mapping in patterns (c) and (d), it is still feasible to produce the mapping between the images containing occlusions by accounting for the mapping that shrinks or enlarges the squares. We also prohibit the mapping in patterns (b), because once these mappings are generated, the mapping around the point indicated by 2 is strongly restricted, thereby leading to a local minimum in the optimization. Consequently, we allow only the mapping consisting of the local transformation in pattern (a). This condition is called the *convexity condition*.

In order to determine mapping which satisfies the convexity condition as a whole, each correspondence in the mapping is determined so that the local convexities are preserved. This can be achieved by restricting the area where the correspondence is sought as follows. When a point $p_{x,y}$ in I_{src} , illustrated by a white point on the left in Fig. 3, is mapped into I_{dst} by a mapping m , the convexity of m is maintained as long as $m(p_{x,y})$ is mapped into the quadrilateral formed by $m(p_{x+\Delta x,y})$, $m(p_{x,y+\Delta y})$, $m(p_{x-\Delta x,y})$ and $m(p_{x,y-\Delta y})$ as illustrated on the right in Fig. 3. Since $m(p_{x,y})$ affects the convexity of only adjacent quadrilaterals, whole convexity of m is guaranteed by beginning with a convex mapping as an initial estimation and then modifying it so that the local convexity is maintained.

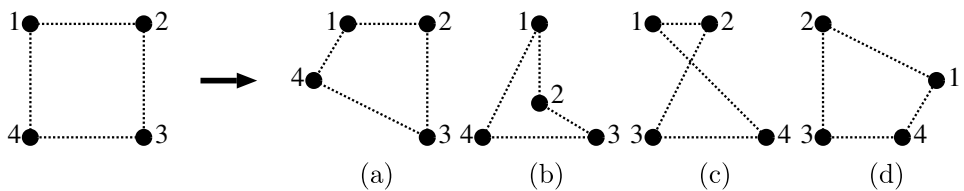


Fig. 2. Possible configurations of four vertices which form a square in one image and are transformed by the estimated mapping. Only pattern (a) is allowed in our method.

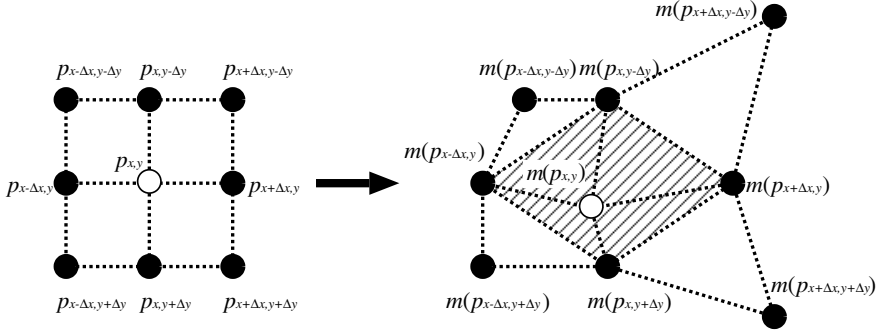


Fig. 3. Search area for single correspondence. The convexity of mapping is maintained as long as each point $m(p_{x,y})$ in a source image (a white point on the left) is mapped into the quadrilateral formed by $m(p_{x+\Delta x,y})$, $m(p_{x,y+\Delta y})$, $m(p_{x-\Delta x,y})$ and $m(p_{x,y-\Delta y})$ as illustrated by a hatched area on the right.

4. Implementation

The algorithm of the plausible image matching can be summarized as follows.

Algorithm 1. (Plausible Image Matching)

Input: Image I_{src}

Input: Image I_{dst}

Output: Mapping M

Local: FilterBank F

Local: FeatureVectorField V_{src}

Local: FeatureVectorField V_{dst}

Local: CostOfMapping c

$M \leftarrow \text{Identity}()$

$F \leftarrow \text{SetupFilterBank}()$

$V_{\text{src}} \leftarrow \text{ExtractFeature}(I_{\text{src}}, F)$

$V_{\text{dst}} \leftarrow \text{ExtractFeature}(I_{\text{dst}}, F)$

repeat

$c \leftarrow \text{CalculateCost}(M)$

for all feature vectors $v_{\text{src}} \in V_{\text{src}}$

$M \leftarrow \text{RefineCorrespondence}(v_{\text{src}}, V_{\text{dst}}, M)$

end for

until $c < \text{Cost}(M)$

The geometrical constraint introduced by the convexity condition makes it difficult to determine all correspondences simultaneously. Instead, we have adopted a sequential method similar to that used in the CPF method.²¹ For each point p in I_{src} , we calculate the cost functions for all possible correspondences which satisfy the convexity condition, and then choose one of the correspondences whose cost is the minimum. The point p to be modified is chosen in a random order to avoid the bias in obtained mapping.

The minimization of the objective function ε^2 is carried out in multiresolutional hierarchy to reduce the risk of the mapping's falling into local minima. Before determining the mapping between images, the feature vector fields generated using the filter bank are repeatedly down-sampled into half the size of the original; that generates pyramids of feature vector fields in different resolutions. We start with the lowest resolution and propagate the calculation toward the highest resolution in turn, using an optimized mapping in a resolution as an initial estimation in the next.

In each resolution in the pyramid, the mapping is refined iteratively until the minimization converges. This iteration is essential in our algorithm to correct erroneous correspondences in lower resolution.

Given image colors as inputs, we apply the filter bank defined in Sec. 3.2 to each color channel in the images separately, and then concatenate obtained feature vector fields into a single field. For instance, if the images are RGB color, we generate the feature vector field composed of 72-dimensional vectors, while gray images yield 24-dimensional vectors.

In our current implementation, the integral in ε^2 is discretized and replaced by a summation for all pixels in source image I_{src} . On the other hand, the domain and range of mapping m is discretized into subpixels. In general, the smaller size of subpixel yields the better mapping, but leads to the larger computational cost. We set the size of subpixels to be 0.1 to 0.5.

5. Experimental Results

In this section, extensive results of the proposed method for plausible image matching are presented. The proposed algorithms are implemented on a standard PC with Pentium3 1 GHz and 768 MB main memory.

5.1. Results of image matching

Figure 4 shows the results of matching between two images of an object which is slightly rotated. The resolution of the image is 128^2 . Figures 4(a) and 4(b) show the input images. Figure 4(c) is the needle diagram for the pixel correspondence determined by the proposed method. Although no geometric information such as camera poses is used in the estimation, the rotation of the object is correctly extracted.

We compared our proposed method with others in terms of quality of mapping generated from the images used in Fig. 4. A comparison of the smoothness of mapping with CPF method²¹ is presented in Fig. 5. The CPF method also imposes a certain geometrical constraint, which the authors refer to as bijectivity condition, on the shape of pixel grids. The essential difference between the bijectivity condition and our convexity constraints is that the former allows the mapping consisting of the local transformation in pattern (b) in Fig. 2. As mentioned in Sec. 3.4, once the

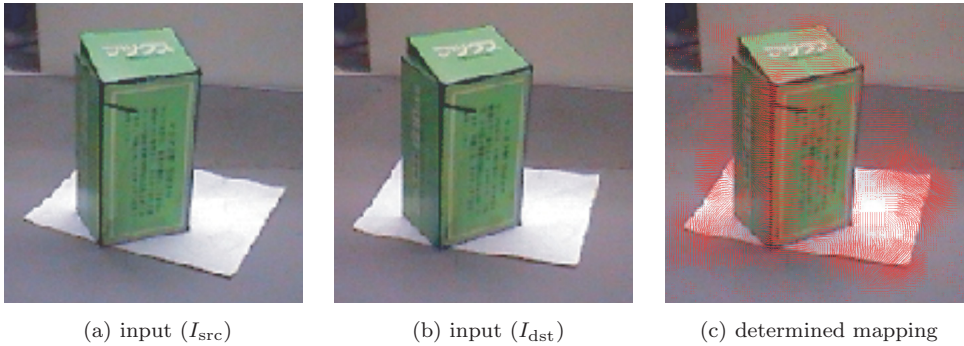


Fig. 4. Example of mapping determined by our algorithm. Given (a) and (b) as inputs, mapping (c) between them is calculated. The mapping is illustrated by needle diagram.

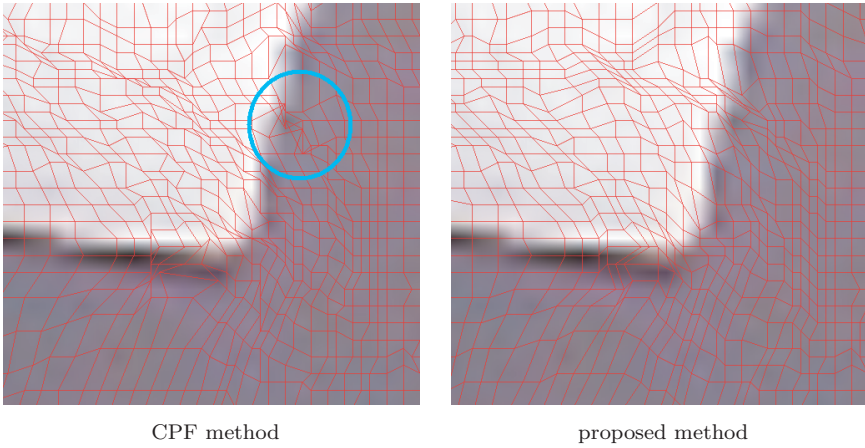


Fig. 5. Comparison of the mapping generated by (left) CPF method and (right) our proposed method. Flipping and twisting of the local correspondence may be observed in the circle on the left.

quadrilaterals in this type of pattern are generated, the mapping around the point is strongly restricted, thereby leading to a strongly distorted mapping. In Fig. 5, some flips and twists of pixels can be observed in the mapping calculated by CPF matching, while our mapping maintains local smoothness.

Figure 6 is a typical example showing the convergence of proposed iterative optimization in multiresolutional hierarchy. The X -axis and Y -axis show the number of iteration and the sum of the squared differences of intensities

$$\varepsilon_I = \sum_{p \in \mathbb{D}} |I_{src}(p) - I_{dst}(m(p))|^2. \tag{6}$$

Each curve in the graph represents the change of ε_I at each level in multiresolutional optimization. ε_I in the graph is properly scaled according to the size of pixel grids

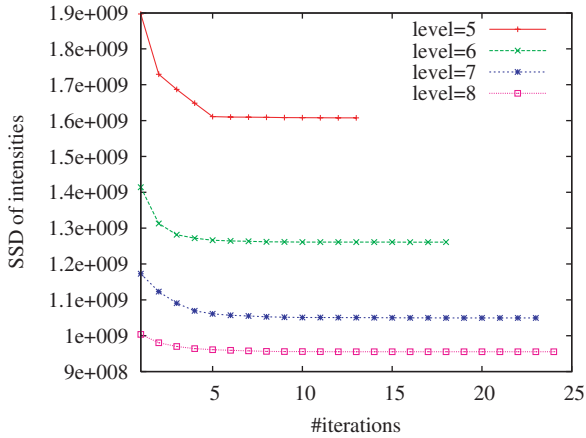


Fig. 6. Convergence of iterative optimization in multiresolutional hierarchy.

Table 1. Performance comparison of image matching algorithms. The left column in each cell shows the execution time (seconds) and the right shows the sum of the squared differences of intensities.

| Data (size) | CPF Method ²¹ | | Proposed Method | |
|---------------------|--------------------------|-------------------|-----------------|-------------------|
| | Time | Error | Time | Error |
| Boxes (128 × 128) | 2.1 | 6.0×10^6 | 8.5 | 5.0×10^6 |
| Textures(128 × 128) | 2.0 | 3.8×10^8 | 10.7 | 1.6×10^8 |
| Faces (256 × 256) | 8.2 | 1.3×10^8 | 51.3 | 1.1×10^8 |

in given images. This graph shows that only a few iterations can drastically improve the quality of image matching.

Table 1 contains a comparison of the matching speed (measured in seconds) and matching quality (measured by sum of the squared differences of intensities) with the CPF method.²¹ The result shows that the quality of mapping has improved, while the speed of calculation has increased by about five times, which depends on data owing to iterative minimization.

5.2. Applications of view-interpolation

Figure 7 shows the result of view-interpolation between two images presented in Figs. 4(a) and 4(b). Although the determined mapping may not be the same as an exact rigid transformation, the in-between views are synthesized without conspicuous visual errors.

Figure 8 shows the result of view-interpolation between two images. The top two figures indicate the reference images for view-interpolation. The synthesized intermediate view-image is presented on the left in the middle row. The mapping between the reference images is estimated in the resolution of 256^2 . The middle right

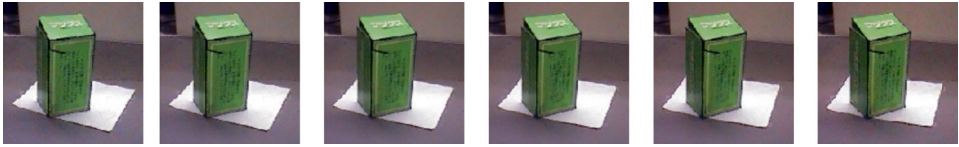


Fig. 7. Results of view-interpolation for a rotated object. This image sequence is generated by trilinear interpolation of the images (a) and (b) in Fig. 4 with automatically estimated correspondence.



Fig. 8. Result of view-interpolation for Pokémons. (a) and (b) are two input images. (c) is the intermediate image synthesized from (a) and (b) by the proposed method of view-interpolation. (d) is a real view-image taken at the viewpoints halfway between those of the inputs and unused in the process of view-interpolation. (e) is the view-interpolation sequence between (a) and (b).

image is the real view-image taken at the position in the middle of the reference viewpoints. Most parts of the synthesized image look quite similar to those of the real image, but small rubber-band effects are observed between green and red toys in the synthesized image. This is caused by interpolating the points which are visible in one of the reference views and invisible in the other due to occlusion. The bottom row presents the sequence of view-interpolation between the given reference view-images.

Figure 9 shows the result of view-interpolation for another scene. This scene has many objects occluding one another; hence, is difficult to estimate accurate mapping. Using the top two images as reference views, the intermediate view-image is synthesized as shown on the left in the middle row. The mapping used to interpolate the views is in the resolution of 256^2 . A view-image taken at the same viewpoint is presented on the right in the middle row for comparison. Owing to the large occlusions and the narrow objects, some parts of the scene correspond to incorrect parts, as shown in the close-up of the synthesized image. The black cable on the orange lamp is doubly observed in the synthetic view as a result of incorrect image correspondences. This effect may be reducible by increasing the resolution of mapping at the cost of computation time.

Image interpolation between the views of a translucent and specular object is presented in Fig. 10. Since the view images are rectangular, they were transformed into squares before determining the correspondences. The mappings between the original images were generated by converting the estimated mapping to that of the original size. It is extremely difficult to reconstruct the accurate geometry of the object, because the change of specular reflection causes misestimation of the correct geometric correspondences between the images. Although the translucency and absence of texture also disturb the estimation, interpolation of the reference view-images enables the system to yield natural synthetic views without any explicit geometric reconstruction.

Note that neither camera poses nor camera intrinsic parameters were estimated in these experiments.

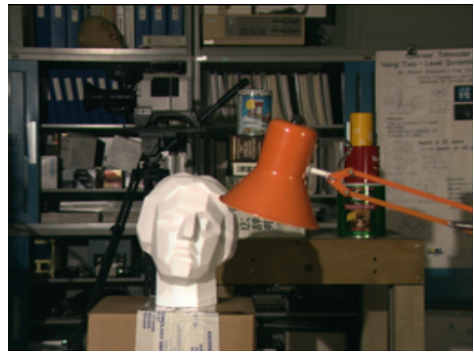
5.3. *Applications of image morphing*

The proposed method of image matching is applicable to the morphing between images of different objects. In Fig. 11, the results of automatic morphing of three different persons are shown. The resolution of the images is 256^2 . The mappings between (a), (c) and (c), (e) are calculated separately. Although the middle person wears the glasses, the robust M-estimator successfully ignores the image features which do not correspond to any part in the counterparts.

Our algorithm was applied to texture metamorphosis as shown in Fig. 12. The size of the texture images is 256^2 . The result is similar in quality to those generated by the semi-automatic method proposed by Liu *et al.*,¹⁴ while our results are automatically generated.



(a) input (I_{src})



(b) input (I_{dst})



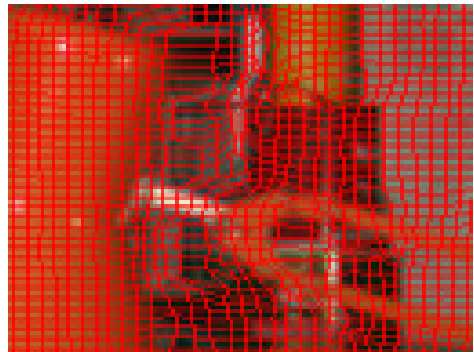
(c) synthesized intermediate view-image



(d) real view-image at the viewpoint of (c)



(e) close-up of (c)



(f) close-up of (c) with grids

Fig. 9. Results of view-interpolation for Tsukuba sequence. (a) and (b) indicate two input images. (c) is the intermediate image synthesized from (a) and (b) by image interpolation. (d) is a real view-image taken at the viewpoints halfway between those of the inputs and unused in the process of the view-interpolation. (e) and (f) are close-up view of the synthesized image. Owing to the occlusion of objects, some parts of mapping are estimated incorrectly.



Fig. 10. Results of view-interpolation for a translucent and specular object. The left-most and right-most images are the reference views and the center is the intermediate image.

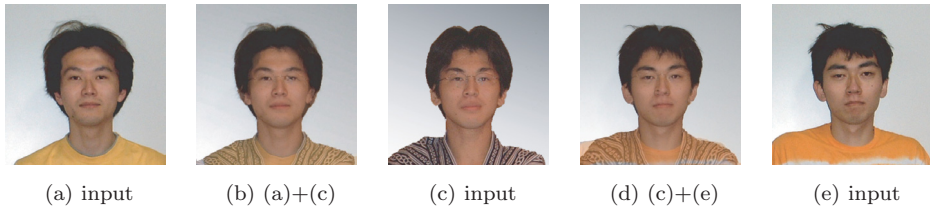


Fig. 11. Blending of human faces. (a), (c) and (e) are inputs. (b) and (d) are generated from adjacent images.

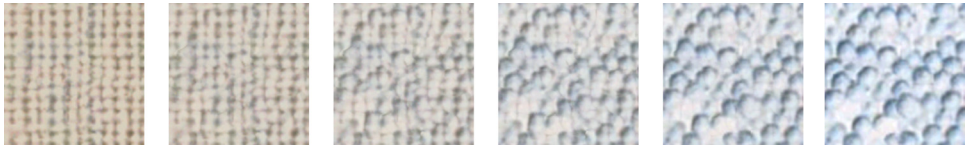


Fig. 12. Texture metamorphosis sequence is automatically generated from left to right.

6. Conclusion

We proposed a method of automatically determining dense and smooth mapping between two images without *a priori* information of either the camera pose or objects. In order to determine plausible correspondences even between different images, features of images are extracted by linear filters similar to those used in early vision. We applied our method to various data sets and showed that our method works better than existing methods when intensity changes or difference between images is large.

Our method tries to find plausible correspondence between given images. However, it is a matter of human perception whether determined correspondence looks natural and plausible. The algorithm will likely fail when given images are considerably different. The physical correctness of correspondence estimated by our proposed method is not guaranteed. The proposed technique is suitable for image-based rendering without explicit geometric reconstruction of the scene which we want to render. An effective algorithm for tuning α^2 and user interface (UI) for user-guided semantic matching remain as future work.

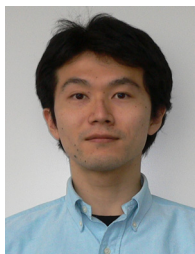
Acknowledgments

The first author's work was supported by grants from RIKEN. The authors are grateful to Dr. Kiwamu Kase for his valuable comments and spiritual support. The authors greatly appreciate Dr. Altion Simo and Dr. Simon Thompson for their English proofreading.

References

1. H. S. Alhichri and M. Kamel, Virtual circles: a new set of features for fast image registration, *Patt. Recogn. Lett.* **24**(9–10) (2003) 1181–1190.
2. R. Bajcsy and S. Kovačič, Multiresolution elastic matching, *Comput. Vis. Graph Imag. Process.* **46**(1) (1989) 1–21.
3. T. Beir and S. Neely, Feature-based image metamorphosis, in *Proc. SIGGRAPH '92*, ACM (July 1992), p. 35.
4. R. Bracewell, *The Fourier Transform and Its Applications*, 2nd edn. (McGraw-Hill, 1965).
5. L. G. Brown, A survey of image registration techniques, *ACM Comput. Surv.* **24**(4) (1992) 325–376.
6. J. Canny, A computational approach to edge detection, *IEEE Trans. Patt. Anal. Mach. Intell.* **8**(6) (1986) 679–698.
7. L. M. G. Fonseca and B. S. Manjunath, Registration techniques for multisensor remotely sensed imagery, *Photogramm. Eng. Rem. Sens.* **62**(9) (1996) 1049–1056.
8. A. Goshtasby, Image registration by local approximation methods, *Imag. Vis. Comput.* **6**(4) (1988) 255–261.
9. K. Habuka and Y. Shinagawa, Image interpolation using enhanced multiresolution critical-point filters, *Int. J. Comput. Vis.* **58**(1) (2004) 19–35.
10. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions* (John Wiley, 1986).
11. B. K. P. Horn and B. G. Schunk, Determining optical flow, *Artif. Intell.* **17** (1981) 185–203.
12. D. Jones and J. Malik, A computational framework for determining stereo correspondence from a set of linear spatial filters, in *Proc. ECCV '92* (1992), pp. 395–410.
13. B. Likar and F. Pernus, Automatic extraction of corresponding points for the registration of medical images, *Med. Phys.* **26**(8) (1999) 1678–1686.
14. Z. Liu, C. Liu, H. Y. Shum and Y. Yu, Pattern-based texture metamorphosis, in *Proc. Pacific Graphics 2002* (2002), pp. 184–191.
15. J. Malik, S. Belongie, J. Shi and T. Leung, Textons, contours and regions: cue integration in image segmentation, in *Proc. Int. Conf. Computer Vision '99* (1999), pp. 918–925.
16. D. N. F. Manfred Ehlers, High-precision geometric correction of airborne remote sensing revisited: the multiquadric interpolation, in *Proc. Image and Signal Processing for Remote Sensing 2315* (1994), pp. 814–824.
17. D. Marr and E. Hildreth, Theory of edge detection, *Roy. Soc. London Series B, Biol. Sci.* **B207** (1980) 187–217.
18. H. S. S. Mike Fornefett and K. Rohr, Elastic registration of medical images using radial basis functions with compact support, in *Proc. Int. Workshop on Biomedical Image Registration* (1999), pp. 173–185.
19. N. R. Pal and S. K. Pal, A review on image segmentation techniques, *Patt. Recogn.* **26**(9) (1993) 1277–1294.

20. W. K. Pratt, *Digital Image Processing*, 2nd edn. (John Wiley, 1991).
 21. Y. Shinagawa and T. L. Kunii, Unconstrained automatic image matching using multi-resolutional critical-point filters, *IEEE Trans. Patt. Anal. Mach. Intell.* **20**(9) (1998) 994–1010.
 22. S. M. Smith and J. M. Brady, SUSAN — a new approach to low level image processing, *Int. J. Comput. Vis.* **23**(1) (1997) 45–78.
 23. G. C. Stockman, S. Kopstein and S. Bennet, Matching images to models for registration and object detection via clustering, *IEEE Trans. Patt. Anal. Mach. Intell.* **4**(3) (1982) 229–241.
 24. G. J. Vanderbrug and A. Rosenfeld, Two-stage template matching, *IEEE Trans. Comput.* **26**(4) (1977) 384–393.
 25. P. A. Viola and W. M. Wells III, Alignment by maximization of mutual information, in *Proc. Int. Conf. Computer Vision '95* (1995), pp. 137–154.
 26. G. Wahba, *Spline Models for Observational Data* (SIAM, 1990).
 27. C.-Y. Wang, H. Sun, S. Uada and A. Rosenfeld, Some experiments in relaxation image matching using corner features, *Patt. Recogn.* **16**(2) (1983) 167–182.
 28. R. A. Young, The Gaussian derivative theory of spatial vision: analysis of cortical cell receptive field weighting profiles, Technical Report GMR 4920, General Motors Research (1985).
 29. Z. Zheng, H. Wang and E. K. Teoh, Analysis of gray level corner detection, *Patt. Recogn. Lett.* **20**(2) (1999) 149–162.
 30. B. Zitova and J. Flusser, Image registration methods: a survey, *Imag Vis. Comput.* **24** (2003) 977–1000.
-



Shuntaro Yamazaki received his B.Sc. (1999) and M.Sc. (2001) degrees in information science, and Ph.D. (2004) in computer science from the University of Tokyo, Japan. He is a full-time researcher at the Digital Human

Research Center in the National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan.

His research interests include computer graphics, vision and its applications. He is a member of the IEEE Computer Society and ACM.



Katsushi Ikeuchi received the B.Eng. degree in mechanical engineering from Kyoto University, Japan, in 1973, and the Ph.D. in information engineering from the University of Tokyo, Japan, in 1978. He is a professor at the Institute of Industrial Science, the University of Tokyo, Japan. After working at the Artificial Intelligence Laboratory at Massachusetts Institute of Technology, the Electrotechnical Laboratory of Ministry of International Trade and Industries, and the School of Computer Science, Carnegie Mellon University, he joined the University of Tokyo, in 1996. He has received various research awards, including the David Marr Prize in computational vision in 1990, and the IEEE R&A K.-S. Fu Memorial Best Transaction Paper award in 1998. In addition, in 1992, his paper, "Numerical Shape from Shading and Occluding Boundaries," was selected as one of the most influential papers to have appeared in the *Artificial Intelligence Journal* within the past 10 years. He is a fellow of the IEEE.



Yoshihisa Shinagawa is currently Associate Professor of Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign. Before joining UIUC, he was Assistant Professor of Department of

Computer Science, University of Tokyo. He received his B.Sc. (in 1987), M.Sc. (in 1990) and D.Sc. (in 1992) degrees in information science from the University of Tokyo.

He is currently Associate Editor of *IEEE Trans. Multimedia*, and is on the Editorial Board of *The Visual Computer* and *The Int. J. Shape Modeling*. He is a member of the IEEE Computer Society and ACM.

His research interests include computer graphics, vision and its applications. He has published more than 74 refereed academic/technical papers in computer science.