Representing an Object by Interchanging What with Where

Jong-Hoon Ahn[1] and Yillbyung Lee
Department of Computer Science, Yonsei University
[1]E-mail:jonghun@postech.ac.kr

Exploring representations is a fundamental step towards understanding vision. The visual system carries two types of information along separate pathways: One is about what it is and the other is about where it is (*1–3*). Initially, the what is represented by a pattern of activity that is distributed across millions of photoreceptors, whereas the where is "implicitly" given as their retinotopic positions. Many computational theories of object recognition rely on such pixel-based representations (*4–9*), but they are insufficient to learn spatial information such as position and size due to the implicit encoding of the where information. Here we try transforming a retinal image of an object into its internal image via interchanging the what with the where, which means that patterns of intensity in internal image describe the spatial information rather than the object information. To be concrete, the retinal image of an object is deformed and turned over into a negative image, in which light areas appear dark and vice versa, and the object's spatial information is quantified with levels of intensity on borders of that image. Interestingly, the inner part excluding the borders of the internal image shows the position and scale invariance (*10, 11*). In order to further understand how the internal image associates the what and where, we examined the internal image of a face which moves or is scaled on the retina. As a result, we found that the internal images form a linear vector space under the object translation and scaling. In conclusion, these results show that the what-where interchangeability might play an important role for organizing those two into internal representation of brain.

The internal image (or conjugate image, see figure 1) explicitly involves the *where* of an object. For instance, figure 2 shows the retinal images of a face and their conjugate images. Even though the face is placed in the left, right, upper, and lower direction, their conjugate images look similar. As the original face converts to its conjugate face, it becomes deformed, turned over, and fitted into the whole screen. Having a good look at the borders of the conjugate images, we can find some different intensity pattern. The pixel intensity on the borders is determined by the spatial locations of a face: The intensity level of the border distant from the face is relatively high. Thus, the conjugate image excluding the borders shows the position-invariant representation.

The inner part of conjugate image explains how the scale invariance can be achieved. Figure 3 demonstrates the scale invariance as well as the position invariance. The position and scale information is incorporated into the intensity on the borders. Thus, the inner parts of three conjugate images in the third column of figure 3 are the same and they are invariant to the position and scale. Figure 3 also provides us with a novel approach to morphological image processing (*12*). By using simple arithmetic operation over the conjugate images, we can effectively control behaviors of the where information. Figure 4 shows the morphological processing of the object images. The lower transition of each example is made by linear combination of two given image vectors, and the immediate images of the transition are the arithmetic result, whereas the upper transitions are regarded as a perceptual mean.

However, the internal image representation cannot be directly applied to multiple or occluded object image. In order to process them, one has to segment a multiple-objects image into single-object images. It would be a good solution to combine internal image representation with layered representation (*13*). The combination can be applied into generating the consecutive images of human walking from limited snapshot images. By means of the representation, we could produce machine-learning based animations.

Another algorithm for producing the automatic morphing transitions was suggested in (*14*). It uses optimal mass transportation theory and concerns the optimal way, in the sense of minimal transportation cost, of moving a pile of soil from one configuration into another. The total amount of mass is required to be constant in the process. By assuming the soil as image intensity, the optimal mass transportation can be applied to automatic image morphing. It matches our method in the aspect that it doesn't require any feature correspondence for morphing. Besides, our formulation can be derived from the optimal transport theory: The optimal transportation from an image to the uniform intensity is the inverse transformation to the optimal transportation from its conjugate image to the uniform intensity, which is depicted in figure 1. The conjugate image is a reciprocal to its original image with regard to optimal mass transportation. However, the automatic morphing methodology via conjugate images gives slightly different results from the morphing algorithm using optimal mass transportation theory; our method is not optimal in the criterion of optimal mass transport, but our method supports a poly-morphing and the series-morphing by adding intermediate images.

Conjugate images depend on the contour of the border. The image boundary that we use is rectangle, and the conjugate faces looks angular and unfamiliar. If the image boundary is circular, the conjugate face will be more globular. And the formalism for conjugate image easily extends to three-dimensional case. The reciprocal relation is still valid: $I(x,y,z)I'(x',y',z')=1$. The range of intensity value $I(x,y)$ includes any positive real number, but we can restrict its lower and upper bound. If the lower and upper bound of $I(x,y)$ is a and b, we can use a scaled intensity $I(x,y)/\sqrt{ab}$ as image intensity: $(I/\sqrt{ab})(I'/\sqrt{ab})=1$. The duality of image representation is compatible with a color representation. The conjugate transformation is applied to each color component such as RGB. If an original image intensity is high, its conjugate image intensity is low due to the reciprocal relation. Thus, the color of conjugate image looks complementary to the original color. For instance, if the color is green, its conjugate color is magenta, whose RGB code is [1 0 1].

The visual system separates the processing of an object's identity ('what') from its spatial location ('where') (*1–3*) and then integrates them in the prefrontal cortex (*15, 16*). But the cortical pathways show appreciable anatomical cross-talk, by which the what and where information are partly interchanged (*17–20*).

Methods Now let us describe a simple mathematics for defining the conjugate image. If we let a continuous positive function $I(x,y)$ be the intensity of a two-dimensional monochromatic image on a confined domain, the intensity of its conjugate image $I'(x',y')$ is defined to be a point-by-point multiplicative inverse of the original image intensity $I(x,y)$ on the domain:

$$I'(x',y') = \frac{1}{I(x,y)}, \tag{1}$$

where the coordinates x and y are irrotationally transformed onto new coordinates x' and y'. The curvilinear coordinates x' and y' are introduced such that the intensity $I(x, y)$ is spread out to have a uniform density in the curvilinear coordinates: $I(x, y)dxdy = dx'dy'$. By the same transformation, the conjugate intensity $I'(x', y')$ is spread out to have the uniform density in the rectangular coordinates as well: $I'(x', y')dx'dy' = dxdy$. As a result, we obtain the reciprocal relation between $I(x, y)$ and $I'(x', y')$ from the two assumptions. The equi-distribution assumption is also shown in figure 1.

The irrotational transformation from (x, y) to (x', y') is characterized as the gradient of a convex scalar function f, that is $x' = \partial f/\partial x$ and $y' = \partial f/\partial y$. Then the Jacobian equation (*21, 22*) becomes the Monge-Ampère equation (*23*): $\det \mathcal{H} f = I$ and $\det \mathcal{H}' f' = I'$, where $\mathcal{H}$ is the Hessian operator. Specifically, when a scalar function is the Legendre conjugate of the other scalar function then their images, or equivalently the determinants of the Hessian of the scalar functions, are defined to be the conjugate of each other as well. Thus, we have to get a numerical solution of the Neumann problem for the Monge-Ampère equation in two dimensions. The procedure consists of three steps: (1) Solve the equation $I = \det \mathcal{H} f$; (2) Derive f' from f; (3) Obtain I' by $I' = \det \mathcal{H}' f'$. Instead of continuous function f, we use discrete expressions f_{ij}, where $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n$. Then the determinant of the Hessian of f is expanded into $\det \mathcal{H} f = m^2 n^2 (F_{11}F_{22} - F_{12}F_{21})$, where $F_{11} = f_{(i+1)j} + f_{(i-1)j} - 2f_{ij}$, $F_{22} = f_{i(j+1)} + f_{i(j-1)} - 2f_{ij}$, and $F_{12} = F_{21} = (f_{(i+1)(j+1)} + f_{(i-1)(j-1)} - f_{(i+1)(j-1)} - f_{(i-1)(j+1)})/4$. The determinant equation of I becomes the quadratic equation with respect to f_{ij}. The equation is analytically solved by fixing all the neighboring f values except for f_{ij}. We have to choose the smaller one between the two solutions because of convexity of f.

Acknowledgements

This work has been supported by the BK21 Research Center for Intelligent Mobile Software at Yonsei University, Korea. This research was also supported as a Brain neuroinformatics Research program by Korean Ministry of Commerce, Industry, and Energy.

References

1. M. Mishkin, L. G. Ungerleider, and K. A. Macko, Object Vision and Spatial Vision: Two Cortical Pathways, *Trends in Neuroscience* **6**, 414-417 (1983).

2. M. A. Goodale and A. D. Milner, Separate Visual Pathways for Perception and Action, *Trends in Neuroscience* **15**, 20-25 (1992).

3. L. G. Ungerleider and J. V. Haxby, 'What' and 'Where' in the human brain, *Current Opinion in Neurobiology* **4**, 157-165 (1994).

4. T. Poggio and S. Edelman, A Network that Learns to Recognize 3D Objects, *Nature* **343**, 263-266 (1990).

5. M. Turk and A. Pentland, Eigenfaces for Recognition, *Journal of Cognitive Neuroscience* **3**, 71-86 (1991).

6. H. Murase and S. K. Nayar, Visual Learning and Recognition of 3D Objects from Appearance, *International Journal of Computer Vision* **14** 5-24 (1995).

7. D. Beymer, T. Poggio, Image Representations for Visual Learning, *Science* **272**, 1905-1909 (1996).

8. M. Pontil and A. Verri, Support Vector Machines for 3D Object Recognition, *PAMI* **20**, 637-646 (1998).

9. H. S. Seung and D. D. Lee, The Manifold Way of Perception, *Science* **29**, 2268-2269 (2000).

10. L. Wiskott, How does our Visual System Achieve Shift and Size Invariance?, *In 23 Problems in Systems Neuroscience, eds. J. L. van Hemmen and T. J. Sejnowski*, publ. Oxford University Press, New York, ISBN-13 978-0-19-514822-0, Chapter 16, 322-340 (2006).

11. X. Miao and R. P. N. Rao, Learning the Lie Groups of Visual Invariance, *Neural Computation* **19**, 2665-2693 (2007).

12. J.-H. Ahn and J.-H. Oh, Unsupervised Morphing by Conjugate Image Processing, *Electronics Letters* **43**, 1137-1138 (2007).

13. M. Pawan Kumar, P.H.S. Torr and A. Zisserman, Learning Layered Motion Segmentations of Video, *ICCV 2005*.

14. L. Zhu, Y. Yang, S. Haker, and A. Tannenbaum, An Image Morphing Technique Based on Optimal Mass Preserving Mapping, *IEEE Transactions on Image Processing*, **16**, 1481-1496, (2007).

15. A. L. Roskies, The Binding Problem, *Neuron* **24**, 7-9 (1999).

16. S. C. Rao, G. Rainer, and E. K. Miller, Integration of What and Where in the Primate Prefrontal Cortex, *Science* **276**, 821-824 (1997).

17. D. J. Felleman and D. C. Van Essen, Distributed Hierarchical Processing in the Primate Cerebral Cortex, *Cerebral Cortex* **1**, 1-47 (1991).

18. D. C. Van Essen, C. H. Anderson and D. J. Felleman, Information Processing in the Primate Visual System: An Integrated Systems Perspective, *Science* **255**, 419-423 (1992).

19. V. P. Ferrera, T. A. Nealey, and J. H. R. Maunsell, Mixed Parvocellular and Magnocellular Geniculate Signals in Visual Area V4, *Nature* **358**, 756-758 (1992).

20. H. Koshino, P. A. Carpenter, T. A. Keller, and M. A. Just, Interactions between the Dorsal and the Ventral Pathways in Mental Rotation: An fMRI Study, *Cognitive, Affective, & Behavioral Neuroscience* **5**, 54-66 (2005).

21. J. Moser, On the volume elements on a manifold, Transactions of the American Mathematical Society 120, 286-294 (1965).

22. B. Dacorogna and J. Moser, On a partial differential equation involving the Jacobian determinant, Annales De L'Institut Henri Poincare 7, 1-26 (1990).

23. D. Gilbarg and N. S. Trudinger, Elliptic Partial Differential Equations of Second Order. Berlin: Springer-Verlag, (1983).

24. A. J. O'Toole, T. Vetterb and V. Blanzb, Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing, *Vision Research* **39**, 3145-3155 (1999).

Figure 1. Conversion of image representation: For the evenly arranged pixels (x, y) and their associated intensity I, this figure shows the conjugate image $I'(x', y')$ by interchanging the pixel positions with the pixel intensity (The abnormal case of unevenly arranged pixels is shown in the supplementary figure 1). Consider a curvilinear transformation $f : (x, y) \rightarrow (x', y')$. By the assumption that its Jacobian determinant J_f is equivalent to the intensity $I(x, y)$ of an image (for instance, the image of a tiger), the what information, which is initially represented by a pattern of intensity, is dissolved into the curvilinear coordinates (x', y'). On the other hand, the even pixel positions are converted into the uniform intensity $I'(x, y)$ by $g : (x, y) \rightarrow I'$, which differentiates the infinitesimal area of a pixel. To complete a duality between original images and conjugate images in a way that the conjugate of the conjugate image is the original image itself, local intensity should be preserved for combining the intensity $I'(x, y)$ and the coordinates (x', y'). Consequently, the conjugate image is equivalent to the inverse Jacobian determinant $I'(x', y') = J_f^{-1}$.

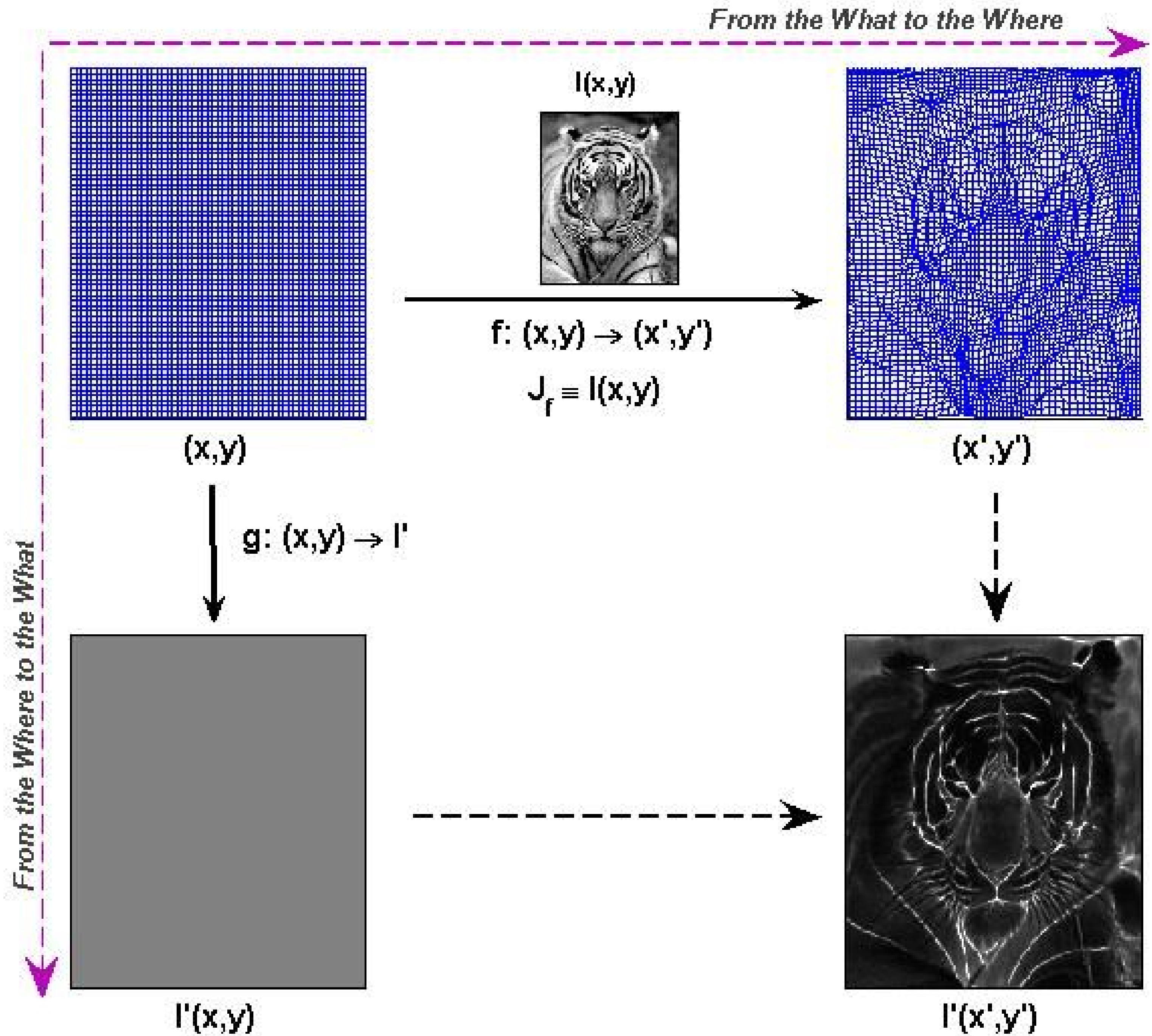
Figure 2. Position-Invariant Representation: The four facial images placed at different positions are transformed into the similarly-looking conjugate images in the third column. But they really have different intensity on the borders. The intensity level of the borders distant from the original faces is relatively high. The inner image excluding the border shows the position-invariant representation or what information under variations in the position of a face on the retina. The facial image is borrowed from (*24*).

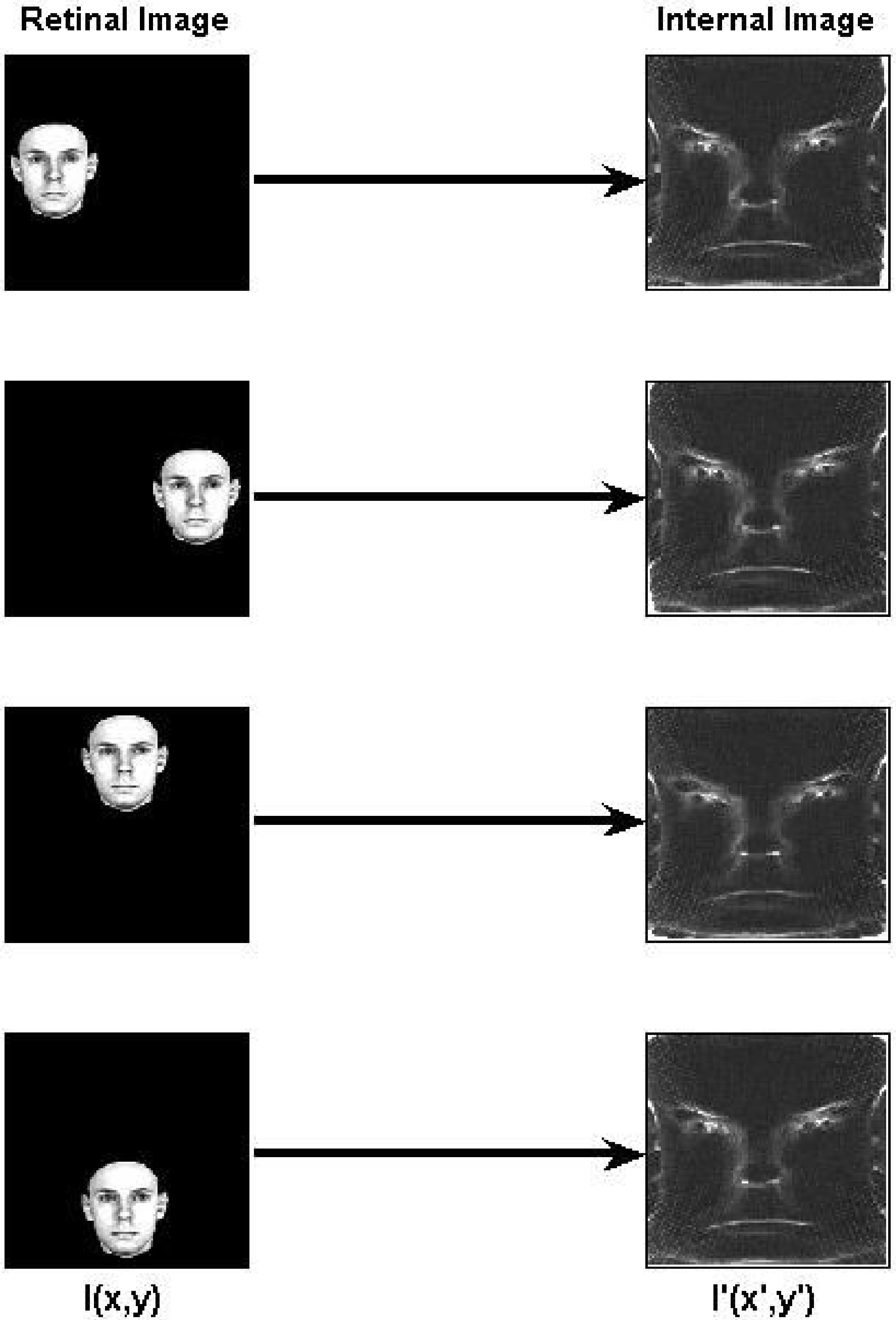
Figure 3. Conjugate Image Processing: Intermediate images are made from two facial images, I_1 and I_2. I_1 is the left-sided small face and I_2 is the right-sided large face. They are transformed into their conjugate images I_1' and I_2' in the third column. The line and arrow represent linear combination and transformation, respectively. Thus the image in the middle of I_1 and I_2 is computed by $(I_1 + I_2)/2$. Similarly, the image in the middle of the conjugate images is computed by $I' = (I_1' + I_2')/2$. The image at the center is again the conjugate of I', which represents a perceptual mean I. This figure shows that the conjugate images form a linear vector space under translation and scaling.

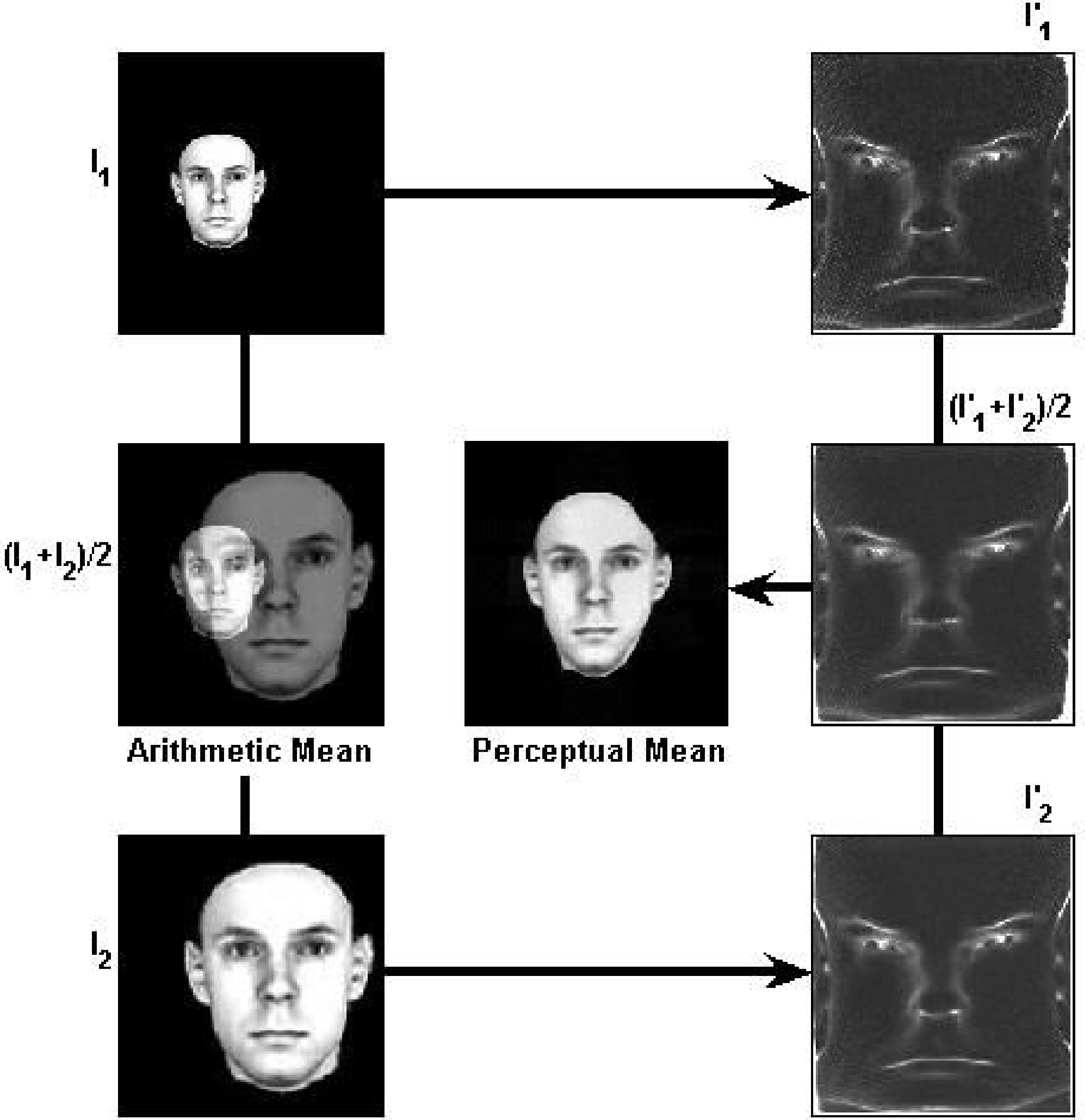
Figure 4. Perceptual Mean versus Arithmetic Mean: We applied conjugate image processing into automatic morphing without any supervision. The upper transitions of each example demonstrate the unsupervised morphing capability from two radically different images: (a)knot and fist; (b)smoke and rose (*12*); (c)sphere and faucet. It is rather difficult to utilize some existing morphing tools, because no strong feature correspondences exist between the images or the smoke shape boundary does not have edges and is vague. Nonetheless, our results show the plausible intermediate images. The lower transitions of each example are obtained by linear combinations of the first and last images. This morphing is useful for images that have such a weak feature correspondence.

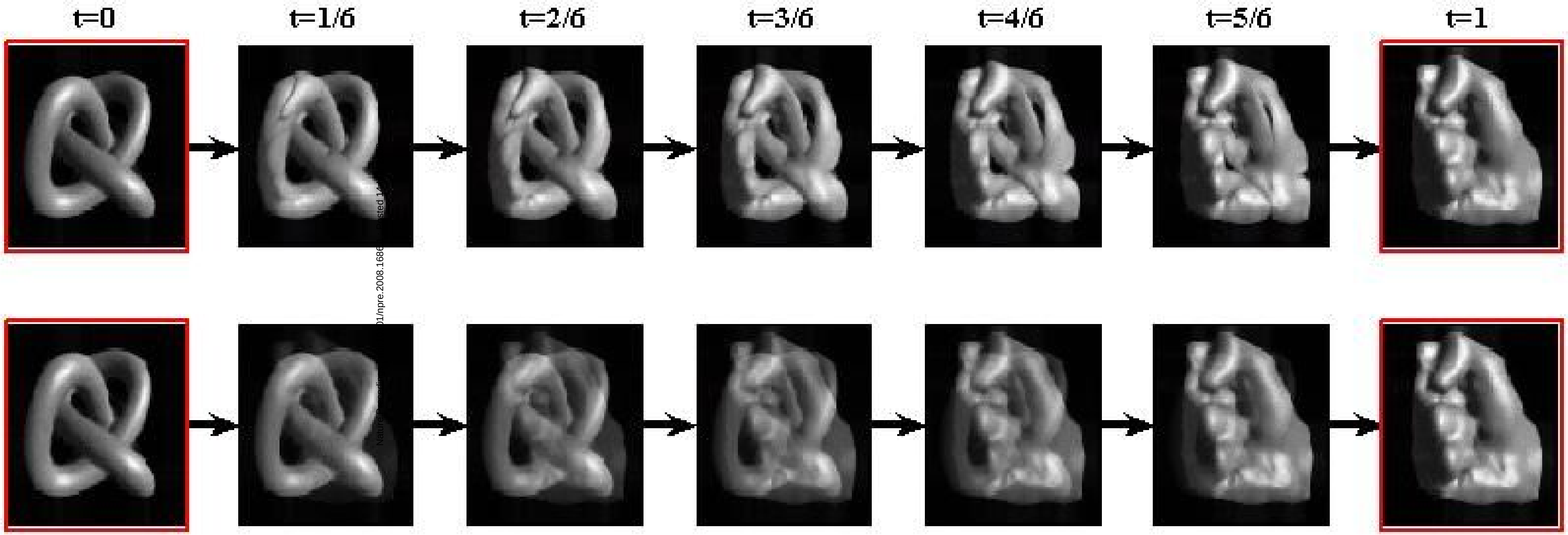
Supplementary Figure 1. Conversion of image representation: For the unevenly arranged pixels (x, y) and their associated intensity I, this figure shows the conjugate image $I'(x', y')$ by interchanging the pixel positions with the pixel intensity. Consider a curvilinear transformation $f : (x, y) \to (x', y')$. By the assumption that its Jacobian determinant J_f is equivalent to the intensity $I(x, y)$ of an image (for instance, the image of a tiger), the what information, which is initially represented by a pattern of intensity, is dissolved into the curvilinear coordinates (x', y'), even if the curvilinear space actually looks flat. On the other hand, the uneven pixel positions are converted into the non-uniform intensity $I'(x, y)$ by $g : (x, y) \to I'$, which differentiates the infinitesimal area of a pixel. To complete a duality between original images and conjugate images in a way that the conjugate of the conjugate image is the original image itself, local intensity should be preserved for combining the intensity $I'(x, y)$ and the coordinates (x', y'). Consequently, the conjugate image is equivalent to the inverse Jacobian determinant $I'(x', y') = J_f^{-1}$.

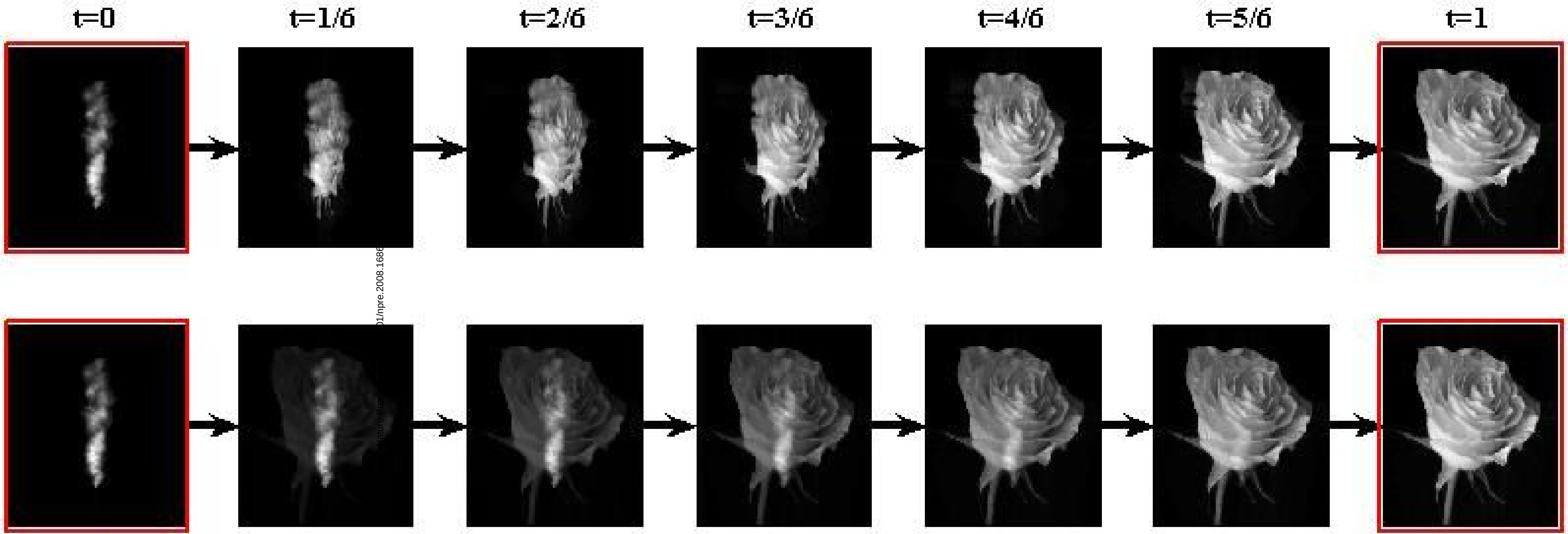
Supplementary Figure 2. Duality of Image Representation: (a)Suppose that there is a volume-preserved elastic slab with a uniform height. When we stretch it horizontally according to a warping map ϕ, the height is changed. The closer the coordinates, the higher the height. Let us consider the height as the image intensity. Then its conjugate image intensity is defined as $|J(\phi)|$, the Jacobian determinant of the warping map ϕ. A warping map describes spatial information of an object image, and its conjugate image intensity or $|J(\phi)|$ tells spatial information rather than object information. (b) The lower images are the conjugate of the upper images. Equivalently, the upper images are also the conjugate of the lower images. As you see the first image pair, the conjugate of a uniform image is the uniform image itself. All the images satisfy the reciprocal relation $I(x, y)I'(x', y') = 1$.

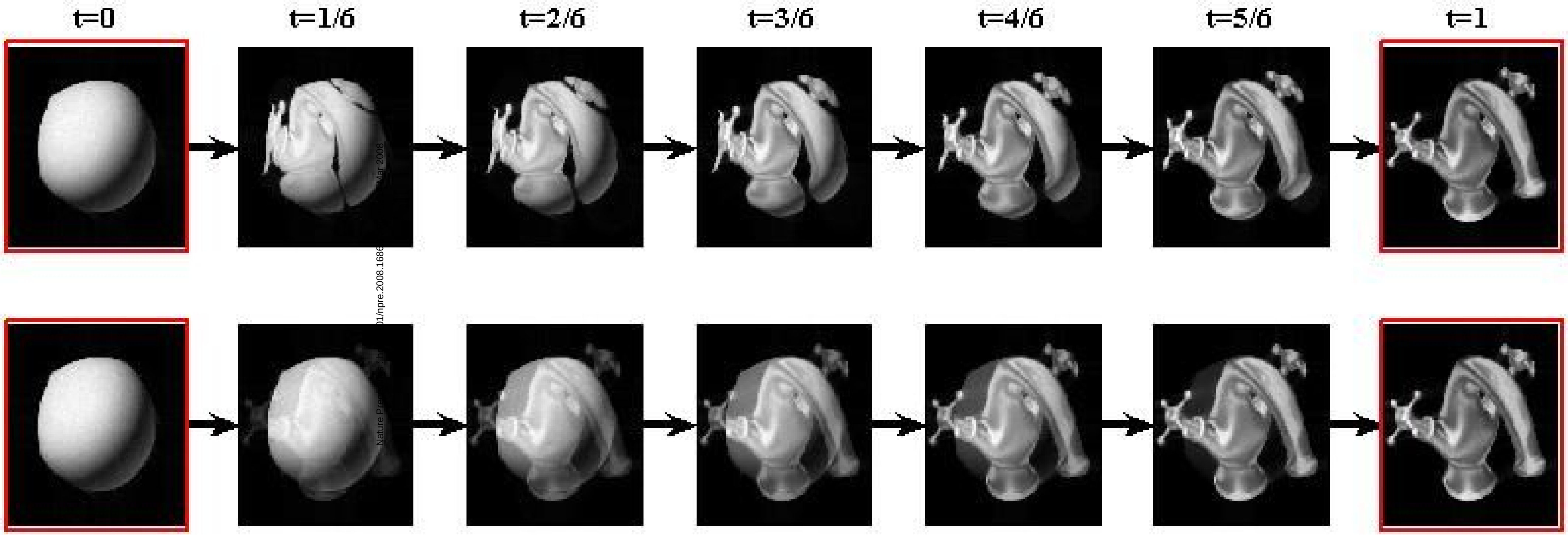

From the What to the Where
From the Where to the What
I(x,y)
f: (x,y) → (x',y')
$J_f \equiv I(x,y)$
(x,y)
(x',y')
g: (x,y) → I'
I'(x,y)
I'(x',y')

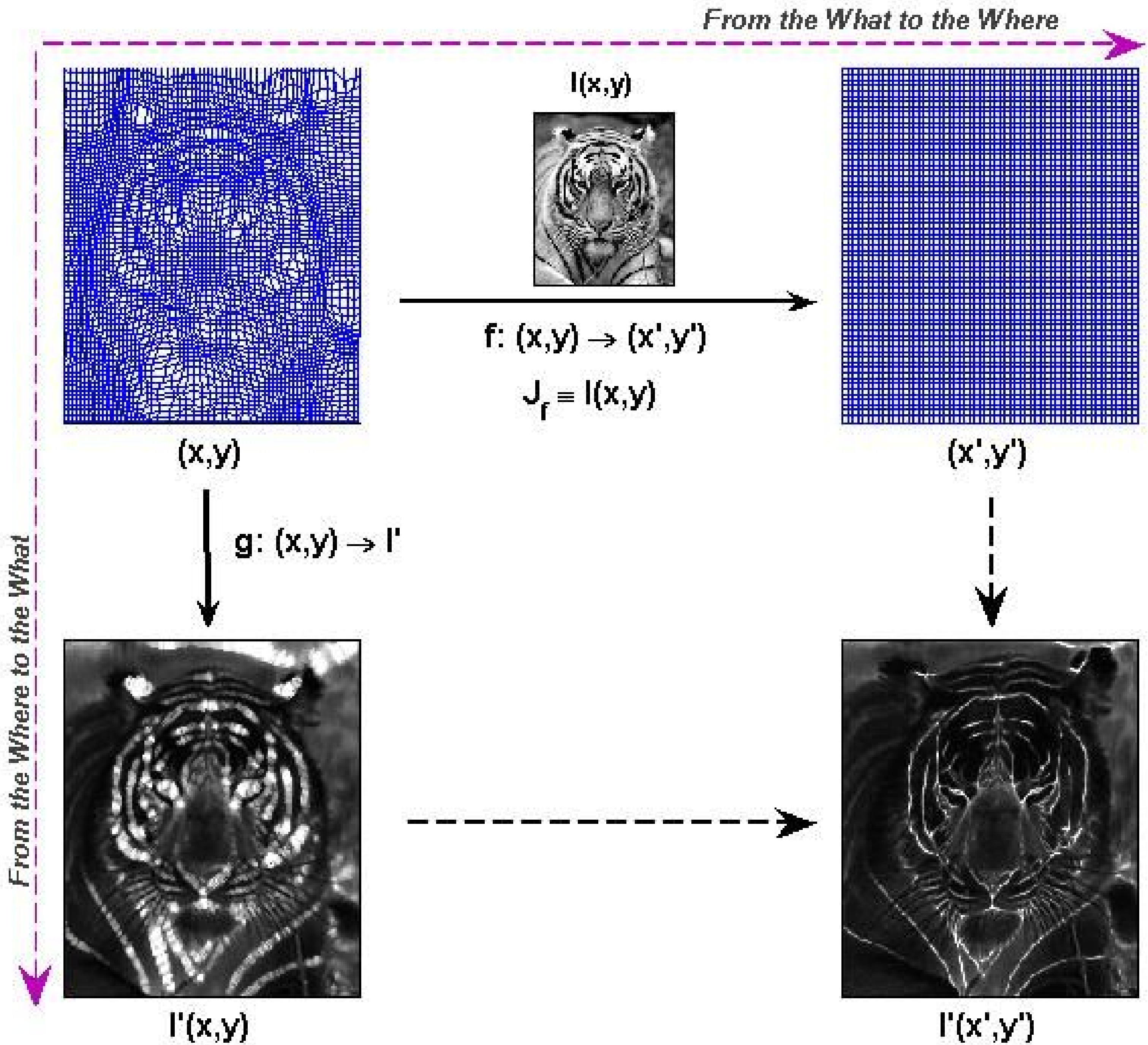

Retinal Image
Internal Image
I(x,y)
I'(x',y')

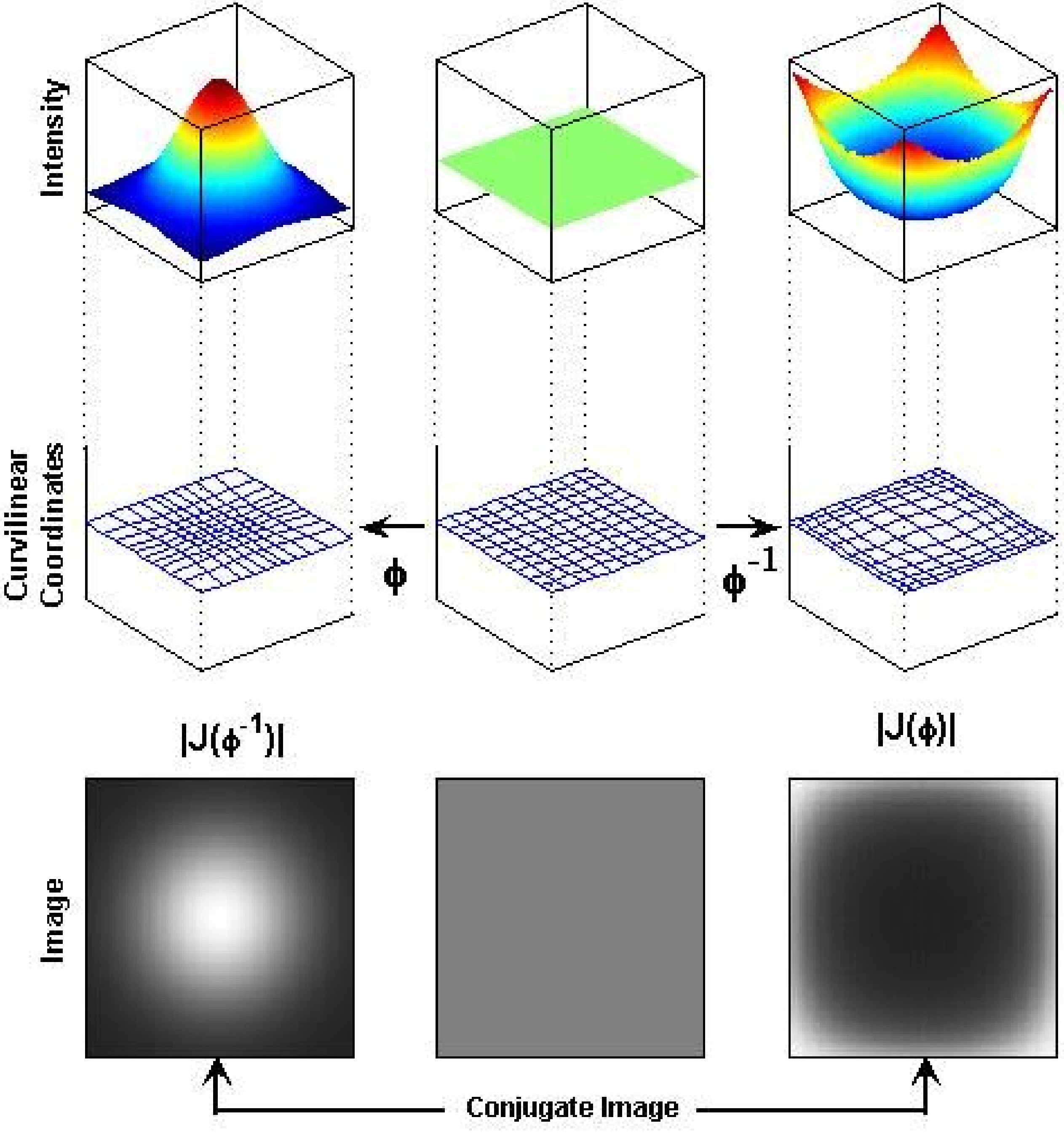

I'1
I1
(I'1+I'2)/2
(I1+I2)/2
Arithmetic Mean
Perceptual Mean
I'2
I2


t=0
t=1/6
t=2/6
t=3/6
t=4/6
t=5/6
t=1


t=0
t=1/6
t=2/6
t=3/6
t=4/6
t=5/6
t=1


t=0
t=1/6
t=2/6
t=3/6
t=4/6
t=5/6
t=1


From the What to the Where
I(x,y)
f: (x,y) → (x',y')
J_f ≡ I(x,y)
(x,y)
(x',y')
g: (x,y) → I'
From the Where to the What
I'(x,y)
I'(x',y')


Intensity
Curvilinear Coordinates
φ
φ-1
|J(φ-1)|
|J(φ)|
Image
Conjugate Image