

SIMILARITY LIMITS OF ORIENTATION CHANGES IN AFFINE TRANSFORMATIONS WITH APPLICATIONS TO PLANAR PATTERN MATCHING

JUAN ZHA¹ AND TIAN XIE^{✉,2}

¹Zhejiang Institute of Mechanical & Electrical Engineering, International Education, Hangzhou, Zhejiang, China; ²Zhejiang Lab, Hangzhou, Zhejiang, China

e-mail: chajuan@zime.edu.cn; rickyskyxie@zhejianglab.com

(Received May 21, 2022; revised June 22, 2022; accepted June 23, 2022)

ABSTRACT

Besides descriptor designing, factoring out the categorical effects of affine transformations is also an effective way to match features. This paper proposed and demonstrated a hypothesis that any coplanar orientations have limited geometrical changes when observed in the camera screen under given ranges of 3D affine transformations. Confirmatory experiments are accomplished by Matlab simulations focusing on specific influences of 3D affine transformations on similarity biases between concerned orientation changes. Statistical analyses and results show that rotations about plane axes mainly affect the geometrical biases of orientation changes. A set of fit limits standing for the biggest influences of most affine transformations in specified ranges are numerically approximated by total sample space distribution and experimental extreme values. We use these Similarity Limits of Orientation Changes (SLOC) as optimization constraints in matching problems, and an application example of planar pattern matching is given. The effectiveness and efficiency of SLOC are proved by the experimental results in Mikolajczyk and HPatches testbed.

Keywords: affine transformation, geometrical orientations, hypothesis testing, planar pattern matching, similarity bias.

INTRODUCTION

Matching is a fundamental problem in computer vision. It plays a deciding role in many applications such like stereo matching (Hamzah *et al.*, 2016), object localization (Ozuysal *et al.*, 2009), image registration (Tomažević *et al.*, 2012), video analysis (Valery *et al.*, 2007), 3D reconstruction (Se and Jaiobedzki, 2006), object tracking (Zhou *et al.*, 2009), and *etc.*. For many instances, the difficulty of matching is caused by the unknown changes of viewpoint which can be explicated as affine transformations in terms of scale, translation and rotation.

To cope with these three kinds of transformations, two trends of approaches have been developed. One is to design the descriptor of the area of interest as invariant as possible to all transformations, represented by famous SIFT (Lowe, 2004), GLOH (Mikolajczyk and Schmid, 2005), BRIEF (Calonder *et al.*, 2010) and FREAK (Alahi *et al.*, 2012). The other is to factor out the categorical effects of transformations, then train classifiers as matching is a classification problem, established by random tree (Lepetit *et al.*, 2005) and FERN (Ozuysal *et al.*, 2007).

The former methods need to compute complex descriptors and measure their similarities at runtime, hence they have a large pressure on ensuring both precision and real-time. The latter relatively reduces runtime computations but whilst needs a high memory cost to represent complicated distributions for every feature, just like the two sides of the same coin. To compensate defects of both methods, some hybrid approaches have been proposed and proved to be efficient.

All these methods have the same task of matching, which is to locate same points in different images. Since points can be connected into determinate lines and lines have determinate orientations in every image, if there exists a similarity of orientation changes of the same line between different images, would the points on lines have some same characteristics under such a geometrical constraint?

In this paper, utilizing scientific method, a similarity of orientation changes is proposed and investigated according to a series of experiments on how affine transformations can affect basic geometrical elements such like angles and lines under the pinhole

camera model, and whether these elements can keep several similarities under simultaneous influences of scale, translation, and rotation.

Results of statistical analysis showed that the similarity of orientation changes is mainly related to rotation transformations, and a set of limits of similarity bias at different ranges of rotation angles are found. These limits stand for the biggest influences of any category of transformations, which is partly following the mentioned scheme that factors out the effects of transformations. But with these limits, classifier would not be necessary for matching. Instead, an optimizer shall be more qualified because this kind of limits can be transformed into some sorts of constraints. The major contribution of our work is a discovery of a novel SLOC constraint which can be applicable for most practical situations.

The rest of paper is organized as follows. First, in material and methods, some basic models are summarized and formulaic conception of proposed geometrical similarity of orientation changes is depicted on emphasis. Then, the design ideas and specific implementation details of experiments are introduced, and interwoven with necessary statistical analyses. Results illustrate the rules and limits found by above experiments and statistical analyses. The mathematical description of deduced constraints are given, we apply these constraints to planar matching strategy as an example. Evaluated in two testbeds, the algorithm can achieve 15FPS for 800×600 image sets and keep precision when facing more rigid conditions comparing to classic matching algorithms, which demonstrates the efficiency of the proposed similarity. Finally, the conclusion and further work are discussed.

MATERIAL AND METHODS

PINHOLE CAMERA MODEL

Matching is to find same points in different images. As points are located in 3D real world objects and imaged to 2D image plane, a mapping from 3D coordinates to 2D image coordinates should be established.

In this paper, we chose the classical pinhole camera model as the basic imaging model because of its commonness. The simple geometry of the camera model is illustrated in Fig.1.

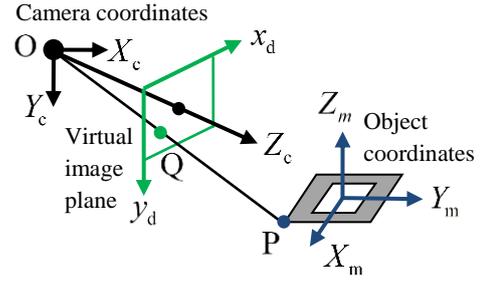


Fig.1. Geometry of the pinhole camera model

Let P be a point in the real-world object at coordinate (X, Y, Z) relative to 3D camera coordinates. Point Q is the projection of point P onto the image plane, at coordinate (x, y) relative to 2D image plane coordinates. Then the mapping from (X, Y, Z) to (x, y) can be given by

$$w \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f \cdot m_x & s & x_0 \\ 0 & f \cdot m_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{P} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (1)$$

where w a scale factor, f the focal length in terms of distance, m_x and m_y the scale factors relating pixels to distance, s the skew coefficient and is set to 0 in our investigation, (x_0, y_0) the principal point and are set ideally in the centre of the image, and finally \mathbf{P} the camera intrinsic matrix which denotes the projective mapping from world coordinates to pixel coordinates in pinhole camera model.

Camera distortions are not considered in this paper for simplicity. Since the procedure of our investigation on geometrical similarity is prone to experimental reasoning, all experiments except application ones are implemented by Matlab simulations in ideal models. Therefore, intrinsic parameters are set to ideal for this virtual imaging system. The influence of distortions could be discussed in the future work if a better precision is to be wanted.

3D AFFINE TRANSFORMATION

3D affine transformation has been widely used in computer vision to describe all ordinary linear transformations such as scale, translation, and rotation. A mapping from coordinate (X, Y, Z) to (X', Y', Z') can be presented as a 9-parameter affine transformation

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \mathbf{WR} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \mathbf{t}, \quad (2)$$

where $\mathbf{W} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix}$ the scale matrix,

$\mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$ the translation matrix and \mathbf{R} the rotation

matrix (Forsyth and Ponce, 2002). \mathbf{R} is generally given by 3 axial rotation α, β, γ .

$\mathbf{R} = \mathbf{R}_1(\alpha)\mathbf{R}_2(\beta)\mathbf{R}_3(\gamma)$ with

$$\mathbf{R}_1(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{bmatrix}, \quad (3)$$

$$\mathbf{R}_2(\beta) = \begin{bmatrix} \cos \beta & 0 & -\sin \beta \\ 0 & 1 & 0 \\ \sin \beta & 0 & \cos \beta \end{bmatrix}, \quad (4)$$

$$\mathbf{R}_3(\gamma) = \begin{bmatrix} \cos \gamma & \sin \gamma & 0 \\ -\sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5)$$

α, β, γ denote Euler angles of x, y, z axial rotations. Every experiment in this paper that involves 3D transformations applies these formulas as basic modes.

SIMILARITY HYPOTHESIS

a. Orientation change

Let l be a straight directed line segment on an object plane and observed by pinhole camera in image I . After some unknown 3D affine transformations, l is observed as l' in image I' . θ, θ' are orientation angles of l and l' in I, I' respectively, representing the orientations of the same line segment in different observation planes, within range of $-\pi$ to π .

Then the Orientation Change (OC) of l from image I to I' with orientations θ, θ' is defined as

$$\text{oc}(\theta, \theta') = \begin{cases} \theta' - \theta & \text{if } \theta' \geq \theta \\ \theta' - \theta + 2\pi & \text{else} \end{cases} \quad (6)$$

with range of $[0, 2\pi)$.

Its geometric significance is the *counterclockwise rotation angle* of the same line's orientation in observation plane. A schematic diagram is showed in Fig.2.

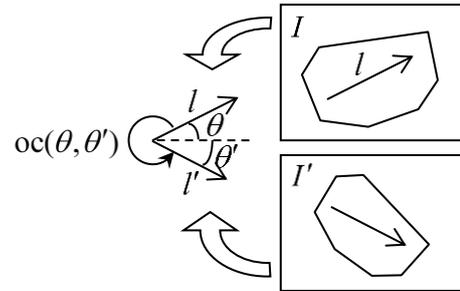


Fig.2 Schematic diagram of orientation change of l between images

In Fig.2, the orientation change $\text{oc}(\theta, \theta')$ is the reflex angle showed in the left. As the orientation angle θ' of l' is less than zero, the orientation change $\theta' - \theta$ has to add 2π to hold positive, just as Eq. (6), to represent the reflex angle.

Furthermore, it is easy to write

$$\text{oc}(\theta', \theta) = 2\pi - \text{oc}(\theta, \theta'), \quad (7)$$

as the reverse OC is the complementary angle of raw OC.

If the directed line segments l and l' have the reverse directions, θ, θ' will add or subtract π separately as they are in range of $-\pi$ to π , and the value of $\theta' - \theta$ will change 0 or $\pm 2\pi$ which can be eliminated by function Eq. (6), as OC is defined in range of $[0, 2\pi)$.

Therefore, the OC is *directionless*. A line segment's OC has no relationship with its direction. We call this line segment the fundamental line segment of its OC.

This definition of OC some terms reflects the influence of 3D affine transformations. Thus our hypothesis can be depicted as "every line segment on the same object plane may have the similar OCs when observed under same conditions".

The suggested hypothesis is a guess that 3D affine transformations may affect OCs in similar ways. We

are trying to find the extent of this similarity by analyzing the extreme of similarities and the distribution of data.

b. Similarity of OC

To test this hypothesis, a simple function Ψ is defined to describe the similarity bias between two OCs:

$$\psi = \Psi(oc_1, oc_2) = oc_2 - oc_1 + 2k\pi \quad (8)$$

where $k \in \{0, -1, 1\}$ is a coefficient to make sure ψ is within the range of $(-\pi, \pi)$. The smaller the absolute value of Ψ is, the more similar two OCs are.

Based on Eq. (8) we can write

$$\begin{aligned} & \Psi(\Psi(oc_1, oc_2), \Psi(oc_1, oc_3)) \\ &= \Psi(oc_1, oc_3) - \Psi(oc_1, oc_2) + 2k\pi \\ &= (oc_3 - oc_1) - (oc_2 - oc_1) + 2m\pi \\ &= oc_3 - oc_2 + 2m\pi \\ &= \Psi(oc_2, oc_3) \end{aligned} \quad (9)$$

where m is a same coefficient like k .

Eq. (9) denotes that the function Ψ is an associative operation just like subtraction, which means the similarity between two OCs can be calculated by utilizing a reference OC. We can first calculate the similarity between the reference OC and either target OC respectively, and then generate the final solution by those two similarities.

This *associative law* is important to be mentioned because it allows us to test any similarity among a large set of OCs by defining only one reference, which largely shortens experimental duration.

CONFIRMATORY EXPERIMENTS

Test pattern

As the similarity function has been defined, the next step is to test the hypothesis by examining ψ under all different conditions refer to two kinds of variables, which are various attitudes of object plane and relative positions of tested line segments.

The attitude of object plane is modeled as a combination of 3D affine transformations in this paper. As mentioned above, these transformations are

categorized into scale, translation and rotation with 9 parameters, and their formulas are given.

The other variable is the position of line segment. As line segments can be distributed in anywhere of the object plane, the distance and intersection angle between them may uncertainly affect the similarity of their OCs. So the incoming experiments should not only test combinations of affine transformations but also the different distribution of line segments.

To satisfy these two demands, we design a 5×5 lattice as the test pattern with every connection of any two vertexes, pictured in Fig.3.

The sampling of line segments distribution can be approximated as an even sampling upon the pattern area, because total 25 vertexes are distributed uniformly in the square pattern, and every vertex has 24 connections to the other vertexes, no less no more. Those 300 connected line segments have total 24 different orientations and 15 different lengths, each orientation has several parallel line segments in different lengths, which can incidentally sampling scale and translation transformations. That is because scale changes only length and distance, and translation alters in parallel. In the pattern there are plenty of parallel lines with various lengths and distances.

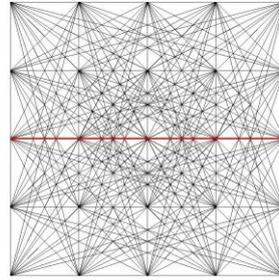


Fig.3 Test pattern. The red color distinguishes the fundamental line of the reference OC.

Therefore, applying the pattern showed in Fig.3 reduces the complexity of confirmatory experiments. The 9 affine parameters need to be tested can be reduced to 3 parameters of axial rotation angles α, β, γ , and the distribution of line segments is also integrated. These two benefits are the major reasons of why we design this pattern.

Experimental design

Our confirmatory experiments are implemented by Matlab simulation in ideal models, so that the camera distortion and physical system errors can be out of consideration.

As mentioned above, the experiments only need to test 3 axial rotation angles (or attitude angles) α, β, γ on x, y, z axis with the proposed pattern. So the task of confirmatory experiments is to examine all ψ between 300 line segments under every variation of these 3 rotation angles. But it's impossible to exhaust all degrees of rotation angles, so we use sampling method again just like the pattern established.

a. Original conditions

Before starting experiments, we first set the original conditions as a reference.

The position of the pattern in camera coordinates is set to $(0, 0, f)$ with attitude angles $\alpha = \beta = \gamma = 0$, where f is the focal length.

To be mentioned, we apply f as the original z axis coordinate is only for simplicity. This value can be changed and the changing can be modeled as a translation on z axis. Since every z axial translation transformation can be replaced by a scale transformation in our camera models, and scale transformations are already sampled by pattern, so we fix the z value as a constant just for the reduction of experimental complexity.

The pattern plane and camera principal plane are parallel. The x axis is horizontal in the pattern and the y axis is vertical. z axis coincides with the camera principal ray. So the original view of the pattern is exactly the view of Fig.3, which is observed as a reference frame or original frame. Thus all OCs are calculated between this original frame and the rotated frame need to be tested.

b. Sampling intervals

The range of α, β is set to $(-90^\circ, 90^\circ)$ in degrees because when rotated by 90 degrees on x or y axis, the 2D pattern can only be seen as a single line in the pinhole camera model. Rotations larger than 90 degrees would just lead to occlusion. The z axis rotations don't have this limitation, so the range of γ is set to $(0^\circ, 360^\circ]$.

The sampling intervals of α, β, γ are all set to 1 degree. Thus there are 11,534,760 ($179 \times 179 \times 360$) different combinations of affine rotation transformations. Consequently, there are 11,534,760 loops of confirmatory experiments need to be executed.

c. Execution details

Between any two different combinations, 300 line segments has 300 OCs, and there are 89,700 (A_{300}^2 , as $\Psi(a, b) \neq \Psi(b, a)$) values of ψ (similarity biases) between these OCs. It is not necessary to calculate all these ψ because we are only interested in distributions and extreme values as mentioned in Eq. (9), when hypothesis is suggested.

Thanks to the associative law of Ψ introduced in Similarity hypothesis, extreme values can be calculated by defining a reference OC and comparing every OC to this reference.

So, we define oc_R as the reference OC corresponds to the horizontal red line in the middle of the pattern (also showed in Fig.3). Then under every different situation of rotations, we just need to compute 300 ψ instead of A_{300}^2 between every OC and oc_R . The calculation can be written as

$$\psi_i = \Psi(oc_R, oc_i) \quad (10)$$

$$\text{where } i \in \{1, 2, \dots, 300\}.$$

Among these 300 ψ , we can find the maximum and minimum value ψ_{\max} and ψ_{\min} , then calculate the extreme value of similarity bias by Eq. (9) as follow:

$$\begin{aligned} \psi_{\text{most}} &= \Psi(oc_{\min}, oc_{\max}) \\ &= \Psi(\Psi(oc_R, oc_{\min}), \Psi(oc_R, oc_{\max})) \\ &= \Psi(\psi_{\min}, \psi_{\max}) \end{aligned} \quad (11)$$

where ψ_{most} stands for the largest bias of similarity between 300 OCs.

To get this ψ_{most} , we record ψ_{\max} and ψ_{\min} in every loop of experiments with their fundamental line segments' positions. The mean value of 300 ψ is also calculated and recorded in every loop marked as ψ_{mean} so is the statistical distribution of 300 ψ which denotes the histogram of these 300 samples with 10 bins.

The experimental results will explain why we record these variables. Statistical analyses of the results will help us discover several laws.

RESULTS

SCALE AND TRANSLATION

Under every combination of rotations, we recorded two extreme values ψ_{\max} and ψ_{\min} , with their fundamental line segments' positions. We draw these "extreme" lines on an empty image which has the same size of the pattern, and superimpose all these images into a single graph, pictured in Fig.4.

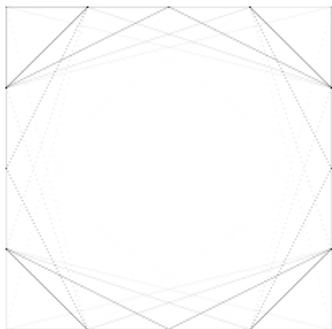


Fig.4 Frequency graph of line segments' positions where holds extreme values of ψ

We call this graph the frequency graph of extreme lines. The lines that have higher gray levels are more frequent to be holding extreme ψ , so we call them extreme lines.

As showed in Fig.4, these extreme lines are located near the pattern edges, which means the further the distance between two line segments, the larger the bias of their OCs while other conditions are same. We call this the *distance law*.

Since the scale transformations are integrated, the middle region of the pattern some terms samples the smaller scales. As extreme lines are only distributed around the boundary, it states that bigger scales bring bigger biases of similarity when other conditions are settled. We call this the *scale law*.

The same is also for translation. As z axial translations can be replaced by scale transformations, we only consider axes parallel to the pattern such as x and y axis. Thus, the extreme line segments around the edges surely have larger translations. So the *translation law* can be illustrated the same as scale, just as bigger translations bring bigger biases of similarity when other conditions are settled.

These laws roughly depict the relationships between similarity biases and scale, translation, or distance inside pattern as they three are potential controlled variables for our experiments. If they are not integrated in our proposed pattern, they should be testified in particular tests merged with rotations. But that will make it too hard to find respective effect of every variables, as they are all fused together. So it is simpler and more effective to do this reduction in our way.

RELATIONSHIP WITH Z DIMENSION

We recorded ψ_{\max} , ψ_{\min} and ψ_{mean} in every loop, and saved them in three $179 \times 179 \times 360$ matrixes as there were equivalent numbers of loops.

All these matrixes are related to x , y , z axial rotation angles, defined as α, β, γ in previous section. So our major task is to find the underlying relationship between various ψ and these three kinds of controlled variables.

Among three axial rotations, only z -axis transformations are parallel to pattern plane. That's the reason why γ has an entire sampling space from 0° to 360° , while α, β are not. So we suppose that the z dimensional rotations may have an independent influence on ψ .

Therefore we extract 2D x - y matrix layers from 3D matrix with every γ , and find that the elements in 2D matrix vary smoothly and rotationally between adjacent layers. They shift in a circle covering the whole matrix, contemporized with the change of γ . This happens the same on matrix Ψ_{\max} , Ψ_{\min} and Ψ_{mean} . We show 6 extracted layers of Ψ_{mean} in terms of depth maps as an instance in Fig.5.

The numbers under every sub image in Fig.5 are layers' indexes corresponding to their z -axis rotation angles γ . Warmer colors stand for the bigger values and cooler ones stand for the smaller. These sub images may not obviously show the smooth changes between layers caused by the big interval, but the trend of anticlockwise rotations can be easily detected. By the way, we don't show the layers after 180 because they are exactly the same as former 180 ones.

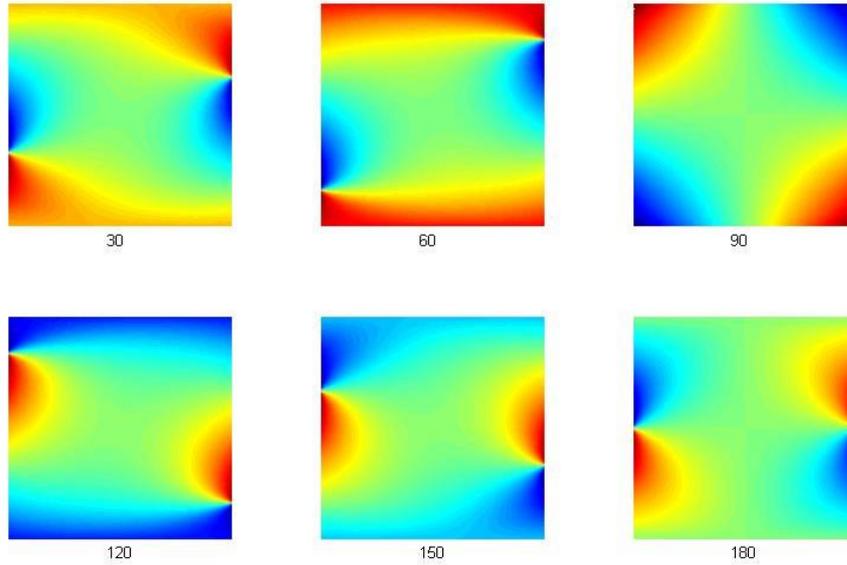


Fig.5 Depth map of extracted Ψ_{mean} matrix layer with different z axis rotation angles

This phenomenon transmits a very important discipline. That is the z axial rotation angles are prone to have little affection on exact values of ψ , but decide the distribution of these values in matrix.

We summarize this law as a result of parallel between pattern plane and rotation orientations. The parallel rotations can only affect the positions of line segments inside the pattern, but the other two orthogonal rotations can conclusively decide the pattern's attitude in real world.

Therefore, next step will focus on how x and y axial rotations determine the similarity bias with less consideration of z dimension.

2D PROJECTION ONTO X-Y

Fig.5 shows the rotational symmetry of our similarity bias matrixes. So we project the matrixes along z dimension onto 2D x - y plane to analyze the control mode of x and y axial rotation transformations.

The projection is not just a simple addition or maximum of values along z dimension, we chose different method for extremum matrixes Ψ_{max} , Ψ_{min} and mean matrix Ψ_{mean} .

The extremum matrix elements are calculated as follow:

$$\Psi_{\text{ext}}(x, y) = \max_{k \in Z} \max \{ |\Psi_{\text{max}}(x, y, k)|, |\Psi_{\text{min}}(x, y, k)| \} \quad (12)$$

where Ψ_{ext} the 2D extremum matrix generated from Ψ_{max} and Ψ_{min} , $|\cdot|$ the absolute function, x and y the any row and column of the matrix, Z the sample space of z dimension, and k the layer index belong to Z .

Eq. (12) expresses the geometrical meaning of Ψ_{ext} as it is the largest absolute bias of similarity under every sampled x and y axial rotation angles.

Every element in Ψ_{max} and Ψ_{min} stands for the extreme values of a 300 ψ dataset comparing to one same reference OC. The maximum of their absolute values can be considered as the biggest geometrical bias among these 300 similarities. So, the maximum in z dimensional sample space expresses the extreme value of ψ that can be caused by every specified attitude of pattern.

Fig.6 (a) shows the 3D shaded surface plot of Ψ_{ext} with its contour plot in Fig.6 (b). The middle of the plot is the origin with no x , y axial rotations, while the boundary area has the biggest rotations.

It is obvious that the values in Ψ_{ext} increase smoothly and radially following the change trends of x , y axial rotations, which clearly expresses the control functionality of x , y axial rotations.

The same can be observed from most bias matrix Ψ_{most} and median matrix Ψ_{med} too. Their formulas are written as follow.

$$\Psi_{\text{most}}(x, y) = \max_{k \in Z} \left\{ \left| \Psi(\Psi_{\text{max}}(x, y, k), \Psi_{\text{min}}(x, y, k)) \right| \right\} \quad (13)$$

$$\Psi_{\text{med}}(x, y) = \frac{\sum_{k=1}^z |\Psi_{\text{mean}}(x, y, k)|}{z} \quad (14)$$

where z is the capacity (360) of sample space Z , and other variables are same as Eq. (12).

The 3D shaded surface plots and contour plots of Ψ_{most} and Ψ_{med} are showed in Fig.6 (c-f) with same scale of axes.

Surfaces showed in Fig.6 are all stalactite shaped. The contours closer to origin are more similar to a circle, while the border ones are more like squares.

The Ψ_{most} has the biggest values of our concerned similarity biases, which is reasonable according to Eq. (11) and Eq. (13).

To be mentioned, Eq. (11) has a necessary condition to hold true, that is the value of $\psi_{\text{max}} - \psi_{\text{min}}$ must be in range of $(-\pi, \pi)$. Otherwise ψ_{most} would add to $2k\pi$ by function Ψ following Eq. (8), thence cannot represent the most bias any more.

The smoothness of values in Ψ_{most} showed in Fig.6 (c) can demonstrate this condition. But the four corner areas in Ψ_{most} are exceptions. We explain this as a result of extreme view angles, under which all points in the pattern are projected onto a single line or a few clustered points, consequently ψ can only be zero or $\pm\pi$, lacking their supposed meanings.

Every element in Ψ_{med} contains an average absolute value of 360 means of every 300 ψ as showed in Eq. (14). The absolute sign is utilized to express the extent of similarity bias, or geometrical bias called in this paper, which is the same in Ψ_{ext} and Ψ_{most} .

Every element in Ψ_{med} contains an average absolute value of 360 means of every 300 ψ as showed in Eq. (14). The absolute sign is utilized to express the extent of similarity bias, or geometrical bias called in this paper, which is the same in Ψ_{ext} and Ψ_{most} .

So, we can conclude that the similarity biases of OCs are mainly controlled by rotations about x and y axis. The bigger the rotation angles, the bigger the bias.

We define this statement as the *rotation law*. It is valid for extreme and mean values according to previous experiments, and is also representing the variation trend of whole sample space in statistics. We then will discuss the distribution of samples, for the purpose of a numerical measure.

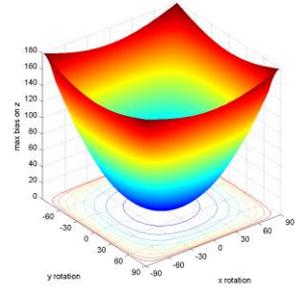
DISTRIBUTION

All the laws found above illustrate several characteristics of extreme values, but not the whole sample space. So we calculate the histogram of all 300 ψ in every loop to learn the distribution of ψ . We replace numerical axis labels of these histograms with abstraction words like ‘min’, ‘mean’ and ‘max’, in order to picture the whole statistical distribution between extremums.

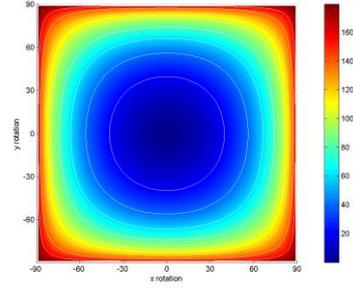
We accumulate all these histograms into a percentage bar graph, which is showed in Fig.7. By simple inspection, three facts can be concluded:

- The distribution is bilateral symmetrical.
- Values close to max or min limits only hold 6.60% of whole sample space.
- The other values approximately follow a uniform distribution.

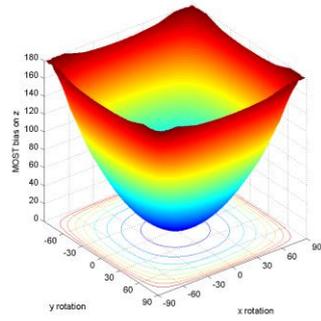
The percentage graph showed in Fig.7 is based on the whole sample space. But every ψ in the space is generated by comparing to the same reference OC not comparing to each other. That means the tested sample space should be considered as a mother space, which gives birth to every possible ψ for any pair of OCs according to Eq. (9).



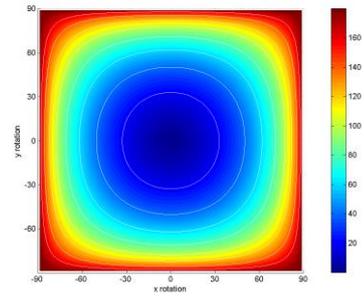
(a) 3D shaded surface plots of Ψ_{ext}



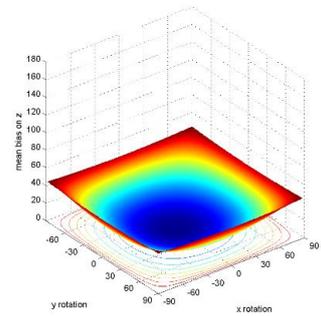
(b) contour plots of Ψ_{ext}



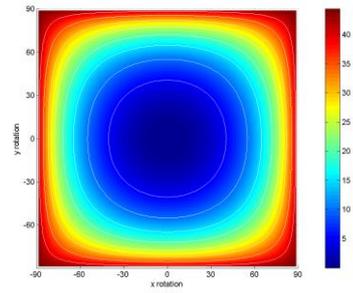
(c) 3D shaded surface plots of Ψ_{most}



(d) contour plots of Ψ_{most}

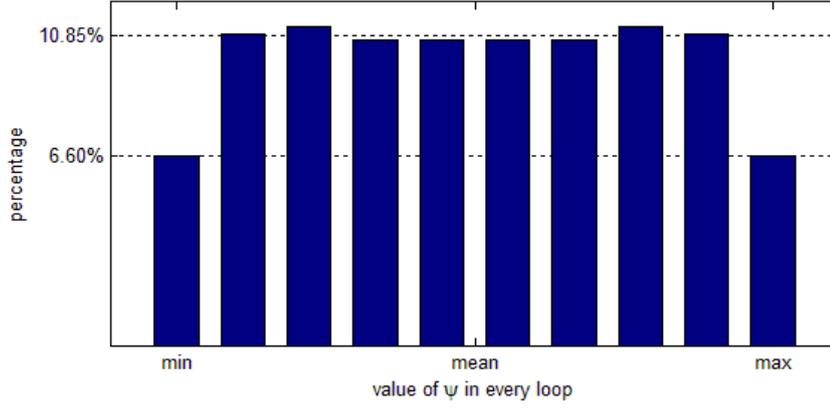
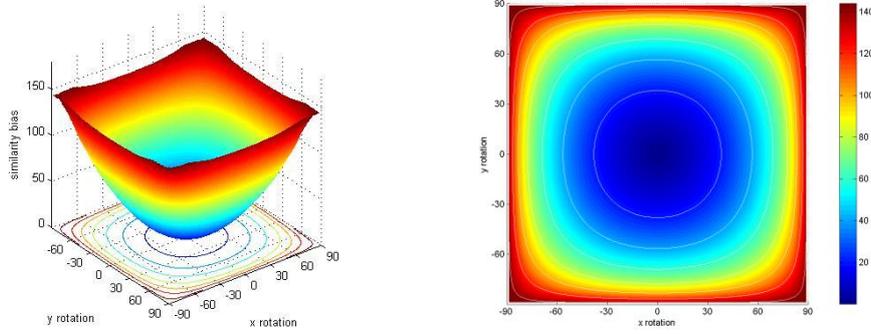


(e) 3D shaded surface plots of Ψ_{med}



(f) contour plots of Ψ_{med}

Fig.6 3D shaded surface plots and contour plots of 2D projected matrixes.


 Fig.7 Percentage bar graph of all ψ , distribution is accumulated in every loop


(a) (b)

 Fig.8 3D shaded surface plots and contour plots of Ψ_{fit}

So, there are 89,700 (A_{300}^2) child pairs of OCs under every rotation, and the largest ones are calculated as Ψ_{most} , which states the limit values of similarity biases. These limits are all calculated by extreme values like ψ_{max} and ψ_{min} in mother space, therefore the largest biases among 89,700 child pairs are also generated by values close to extremums in mother space, which occupy 13.2% of whole quantity and 20% of value bins round border as showed in Fig.7.

Thus, the largest biases have approximately 0.87% ($6.6\% \times 6.6\% \times 2$) of quantity in child space. Which means the other 99.13% of biases are all below 80% of the value limit ψ_{most} according to following equations.

$$\left. \begin{aligned}
 \psi_{most} &= \psi_{max} - \psi_{min} \\
 \psi'_{max} &= (\psi_{max} - \psi_{mean}) \cdot 80\% + \psi_{mean} \\
 \psi'_{min} &= (\psi_{min} - \psi_{mean}) \cdot 80\% + \psi_{mean}
 \end{aligned} \right\} \Rightarrow$$

$$\begin{aligned}
 \Rightarrow \psi'_{most} &= \psi'_{max} - \psi'_{min} \\
 &= (\psi_{max} - \psi_{min}) \cdot 80\% \\
 &= 0.8\psi_{most}
 \end{aligned} \tag{15}$$

where ψ'_{max} and ψ'_{min} stand for the new limit values in mother space when cutting off the extremum bins at two ends of percentage graph in Fig.7. ψ'_{most} stands for the new biggest bias. $2k\pi$ has not been added for ψ_{most} and ψ'_{most} in Eq. (15) because it has been discussed when Ψ_{most} was first introduced.

The number “80%” is came from the bin numbers in Fig.7, as value’s range is divided into ten equal parts, and bilateral symmetrical with mean value.

Then, following Eq. (15) we can use the known Ψ_{most} to estimate the numerical limit of the majority’s similarity biases. We call this limit the “fit limit” Ψ_{fit} . Fig.8 shows the 3D shaded surface plots and contour plots of Ψ_{fit} , which are all 20% smaller than Ψ_{most} .

FIT LIMITS

a. Limit values

Fig.9 draws the numerical comparison of Ψ_{most} and Ψ_{fit} in 2D plots when x or y axial rotation is zero. The red line showed in Fig.9 stands for the one-dimensional data in 2D matrix Ψ_{most} when x or y rotation value equals to zero, and the blue line stands for the same in Ψ_{fit} . The blue line is 20% lower than red line just following the Eq. (15).

So, we can get the numerical fit limit at different ranges of x and y rotations according to Fig.9. We represent them in Table 1, where α, β are x and y axial rotation angles, and limits of similarity bias are the ceiling function (smallest following integer) of true values on the blue line in Fig.9, with true values showed in parentheses.

We use that ceiling function for two reasons. First, we use it to tolerate the errors came from undefined display precision, as the calculation of OC is dependent uniquely on pixel positions in camera observation plane. Which means higher precision images may have sharper line segments, and consequently have better accuracy on calculations of OCs and ψ .

In our experiments, we utilize a specified precision of 800×600 pixels for observation image. So, in real applications, any higher or lower precision would generate errors. But these errors wouldn’t be very large, as display precision can only slightly affect the orientations of lines.

The second use of ceiling function is to roughly tolerate the unmodeled camera distortion and possible systematic errors of real applications. However, the exact influences of these errors are not measured in this paper. Thus, in this tentative usage, ceiling function is just a symbolical compensation, or a subjective guess for toleration.

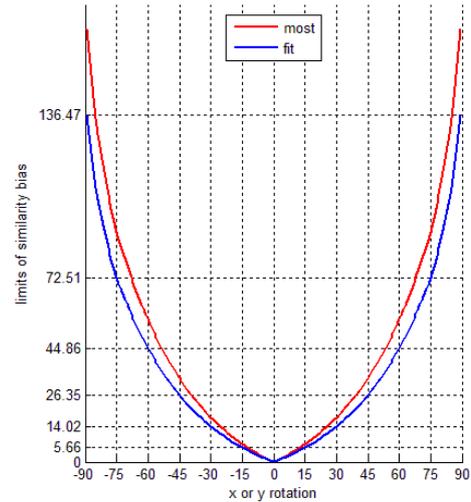


Fig.9 2D plots of Ψ_{most} and Ψ_{fit} when x or y axial rotation is zero

Table 1 Fit limit for different ranges of rotations

$\alpha, \beta \in$	Limit of similarity bias
$(-15^\circ, 15^\circ)$	6° (5.66)
$(-30^\circ, 30^\circ)$	15° (14.02)
$(-45^\circ, 45^\circ)$	27° (26.35)
$(-60^\circ, 60^\circ)$	45° (44.86)
$(-75^\circ, 75^\circ)$	73° (72.51)
$(-90^\circ, 90^\circ)$	137° (136.47)

b. Ranges

After explaining how and why we handled limit values in Table 1, it is equally important to describe the actual meaning of corresponding ranges of α and β in the left column.

Our true intension is to find the fit limit of similarity bias in reasonable ranges. So, we samples the x and y axial rotation angles to check different observation view of the pattern. Fig.10 pictures several samples on individual rotations about y axis. It can be seen that when rotation is bigger than 60° , the pattern is so slant that it can only be observed in half size or even narrower. Then lines in pattern would be squeezed into a gathered bunch, hereupon, too close to be distinguished.

Therefore, we believe that $(-60^\circ, 60^\circ)$ is a reasonable range with certain application values, as it

covers most applications' perspective conditions with few harsh requirements on line segments' detection.

But the range in terms of $\alpha, \beta \in (-\theta, \theta)$ doesn't mean an ordinary square region in x - y coordinate plane. It actually stands for a contour area that holds the exact limit value of corresponding fit limits listed in Table 1.

As showed in Fig.8, the contours are horizontally and vertically symmetrical. Their shapes are prone to be circles when near the center of coordinates, and squares when near the edges. It is hard to fit all these contours with mathematical formulas, so we just use simple notations like $\alpha, \beta \in (-\sigma, \sigma)$ to approximate the axial borders of contours' range area.

c. Statement

Now we can conclude that in a rotation range of approximate $(-60^\circ, 60^\circ)$ about x and y axis, the absolute bias of similarity between any two orientation changes of coplanar line segments is always less than 45° under arbitrary 3D affine transformations including scale, translation and rotation. The smaller the rotation range, the smaller the corresponding limit, which just follows the Table 1.

We call this conclusion the Similarity Limit of Orientation Changes (SLOC). It is the numerical verification of our hypothesis. As it gives the exact description of similarity limitation and applicable situations, we can apply it into planar pattern matching as a geometrical constraint in order to find the same features in different images.

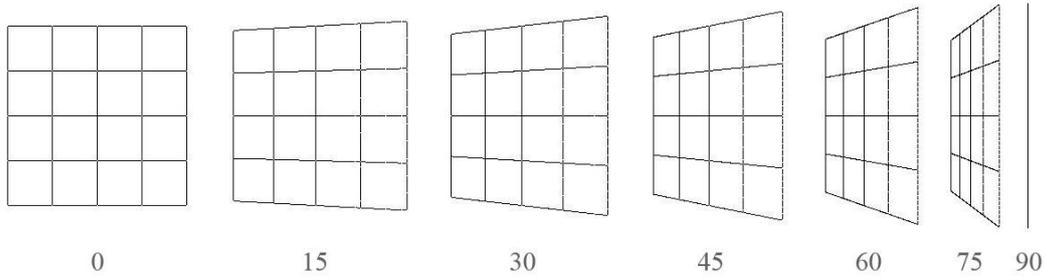


Fig.10 Observation views under different rotation angles (in degree) about y axis

APPLICATION TO PLANAR PATTERN MATCHING

Our SLOC primarily focuses on the lines in planar pattern. It can also be applied to match points or regions as we can draw auxiliary lines between points or regions.

In this section, we will depict an example on planar pattern matching to illustrate the application methods of SLOC.

The example is based on a previous work of augmented reality markerless registration (Xie *et al.*, 2013). In that work, we took colorful connected areas as invariant features, calculated their descriptors simply by hue and geometry information, and matched them by global optimization based on geometric constraints under undefined view transformations.

Among those geometric constraints, the major one is the SLOC which was proposed and applied as an

empirical constant at that time and is formally studied and testified in this paper.

Let's simply review the primary strategy of the algorithm and illustrate the usage of SLOC in matching points or region features.

a. Algorithm strategy

Inspired by the visual attraction of various colors in nature worlds, we apply regions with the uniform hue as invariant features because Hue is invariant to illumination changes and view transformations.

Classical flood-fill method is utilized to achieve feature extraction. The description of these region features is composed of region's average hue, contour shape, center point, area size, and region orientation.

Average hue is the most basic matching criterion. As we intend to find same region features under different views, we firstly filter out regions whose average hue is far different from reference ones.

Then we calculate Hu invariant moments (Hu, 1962) on candidate regions to compare their contour shapes and filter out unmatched ones again, as the second step.

Thus, for every region feature in the reference frame or pattern, we can obtain a candidate matching list. Which makes the matching problem become finding the global optimal solutions among these candidate lists.

We accomplish this optimization in an iterative way similar to RANSAC (Fischler and Bolles, 1981). In every iteration, we pick one region feature as a seed region, and assume one of its candidate matches on the list is a correct matching, called the seed matching. Then we examine every other region features by this seed matching with several geometrical constraints to find at most one (could be zero) best match for each tested region features.

Then in one iteration we can get one global solution of pattern matching with an independent list of matched features. In different iterations the number of matched features may not be the same. So, the final global solution is the one has the longest list of matched features, or the one has the highest score of matching quality when several lists hold the same length.

This optimization strategy is simple and efficient. It applies SLOC as the major constraint since there are two kinds of orientations related to regions between frames, and our SLOC is basically handling changes of orientations.

b. Orientations

Fig.11 illustrates the mentioned two kinds of orientations. They are the orientation of region, symbolized by θ , and the orientation between center points, symbolized by φ .

In Fig.11, the blue rectangular region R_{seed}^{ref} and red elliptical region R_{test}^{ref} are the seed and test region features in the reference frame. The similar-shaped region R_{seed}^{cur} and R_{test}^{cur} in the current frame are the candidate matches of R_{seed}^{ref} and R_{test}^{ref} respectively.

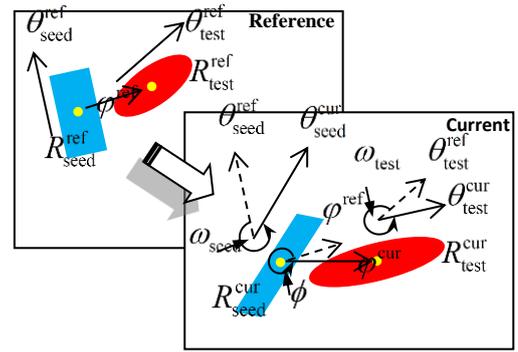


Fig.11 Orientations of interest between frames

θ_{seed}^{ref} and θ_{seed}^{cur} are the orientations of region R_{seed}^{ref} and R_{seed}^{cur} in the reference frame and current frame respectively. They are defined by the Principal Component Analysis (PCA) of points' distribution inside regions. The same are the θ_{test}^{ref} and θ_{test}^{cur} for the test region.

If we assume R_{seed}^{ref} and R_{seed}^{cur} are correctly matched as the same region feature, the OC (orientation change) of this seed region can be generated by θ_{seed}^{ref} and θ_{seed}^{cur} as showed in Fig.11, with the symbol ω_{seed} . The same of ω_{test} can also be seen in Fig.11 for the test region.

φ^{ref} is the orientation between center points of seed region and test region in the reference frame. It depicts the directions from seed region to test region. So is the φ^{cur} in the current frame. Their OC is symbolized as ϕ .

According to SLOC, the absolute biases of similarity among these three OCs should be less than 45 degrees. So, we can write following equations.

$$|\Psi(\omega_{seed}, \omega_{test})| < 45^\circ \quad (16)$$

$$|\Psi(\omega_{seed}, \phi)| < 45^\circ \quad (17)$$

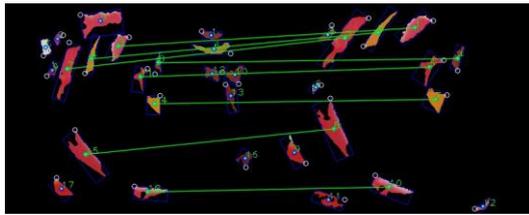
Eq. (16-17) are the major constraints on filtering out candidate matches. The change of area size and distance between regions are also considered as an addition. When test region has more than one suitable match, the best match would be the one with the smallest similarity bias.

Now we have described the example usage of SLOC studied in this paper, we will then examine the precision and efficiency of this usage by experiments.

c. Experimental results

Mikolajczyk testbed (Mikolajczyk *et al.*, 2005) and HPatches (Balntas *et al.*, 2017) are applied to evaluate the capability of the above registration algorithm following SLOC constraint. The standard ‘Graffiti’ scene in Mikolajczyk testbed and ‘Wormhole’ scene in HPatches are applied as they contain various colors and big viewpoint changes.

Fig.12 depicts an example of the optimization result of colorful regions matching. It demonstrates the effectiveness of SLOC.



(a) Region matching



(b) Matched regions in the scene

Fig.12 Example of matching result

Fig.13 depicts the final registration results of graffiti and wormhole scene, with the comparison to classic local descriptor SURF (Bay *et al.*, 2008), ORB (Rublee *et al.* 2011), and BRISK (Leutenegger *et al.*, 2011) methods. The upper row images depicted in Fig.13 are the registration results of graffiti scene in more and more tilt viewpoints and the lower row corresponds to wormhole scene respectively. The first image in each row is the reference image to be registered.

The wireframes in different colors drawn in Fig.13 (b-e, g-j) are the registration results of different algorithms compared in this paper. The color of yellow, green, blue, purple, and magenta respectively stand for the ground truth of the testbed and the result of SLOC, SURF, ORB, BRISK. The errors of each registration result are estimated by the average distance of four corners of wireframes between each tested algorithm and ground truth (Lieberknecht *et al.*, 2009).

Table 2 and Table 3 depict the numerical measure of registration errors and matching speed of our SLOC algorithm and SURF, ORB, BRISK. All experiments are implemented by the hardware of Intel (R) Core (TM) i7-10750H Dual-CPU @ 2.60GHz 2.59GHz, with 64GB memory, while software platform is OpenCV 2.4.13.6 with integrated SURF, ORB, BRISK algorithms under 64-bit Windows 10 operation system.

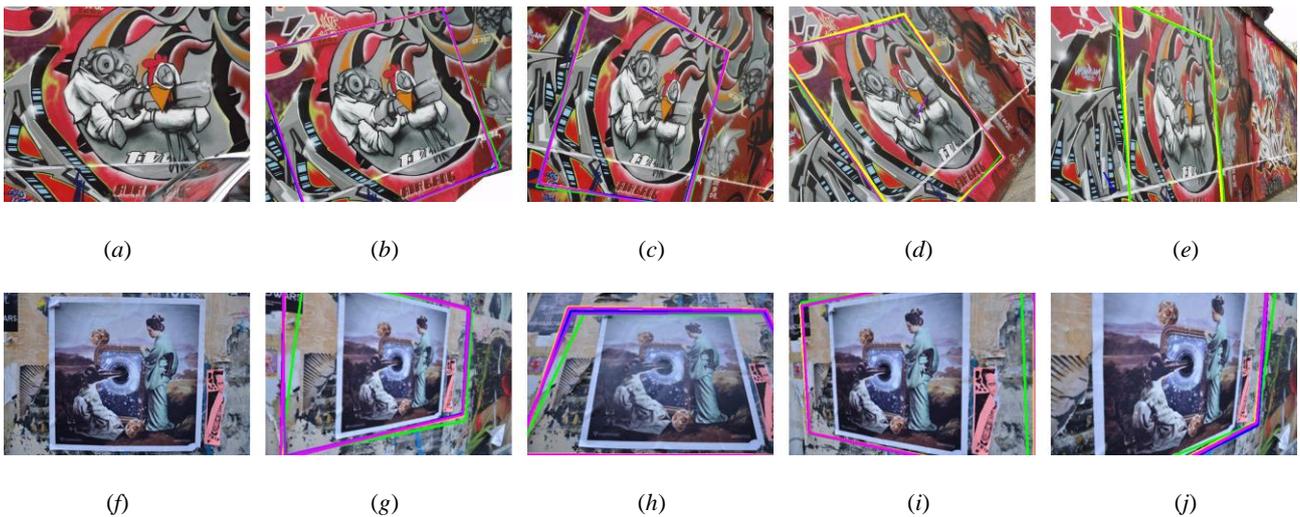


Fig.13 Registration results with comparison to SURF, ORB, BRISK. Upper row: graffiti scene of Mikolajczyk testbed. Lower row: wormhole scene of HPatches testbed. The yellow, green, blue, purple, and magenta wireframes stand for the registration results of given ground truth, SLOC, SURF, ORB, and BRISK respectively.

Table 2 Comparison of registration errors (pixels)

Sub-Fig.	(b)	(c)	(d)	(e)	(g)	(h)	(i)	(j)
SLOC	5.86	11.46	21.28	14.63	17.11	15.37	13.52	43.58
SURF	3.93	13.07	518.17	514.83	3.41	6.31	2.09	13.08
ORB	1.96	4.61	406.89	421.25	3.46	13.95	3.52	7.87
BRISK	1.35	8.66	11.73	507.74	2.75	1.15	2.16	3.49

Table 3 Comparison of average speed (ms)

Testbed	Graffiti (800×600)			Wormhole (493×327)		
	Feature detection & descriptor extraction	Matching	Average FPS	Feature detection & descriptor extraction	Matching	Average FPS
SLOC	45.53	17.02	15.98	42.51	2.35	22.58
SURF	423.48	113.28	1.86	112.97	12.21	8.06
ORB	139.76	3.58	6.99	60.79	3.75	15.5
BRISK	336.75	123.32	2.18	95.48	10.45	9.44

Depicted in Table 2, the registration results of SLOC are not better than classic algorithms when the tilt of viewpoint is relatively small. However, when the tilt is getting larger such as in Fig.13 (d-e) sub-images, classic algorithms almost fail while SLOC still keeps good precisions. This evidence demonstrates the capability of our SLOC algorithm under large viewpoint changes. It's also a practical testification as our SLOC is theoretically suitable for tilt angles below 60 degrees.

On the other hand, Table 3 illustrates the efficiency of our matching strategy compared to classic algorithms. As showed in Table 3, our algorithm can achieve 15 FPS real-time matching for 800×600 graffiti scene while classic algorithms can only achieve less than 7 FPS.

All in all, the algorithm example described in this section can be thought as an instruction of our SLOC constraint for planar pattern matching. The experimental results showed both good effectiveness and efficiency. There could also be many other ways to utilize the SLOC when facing other kinds of conditions. We shall keep on studying the application method of this category of matching constraints, and other constraints like distance and area size should be developed too.

DISCUSSION

In this paper, with the help of hypothesis testing, we find a series of Similarity Limits of Orientation Changes (SLOC) under specified range of 3D affine transformations.

The mathematical expressions of Orientation Change (OC) and the similarity bias (Ψ) between orientation changes are proposed. Confirmatory experiments are designed to study the specific influence of 3D affine transformations on these similarity biases. A lattice test pattern integrated with translation and scale simulations is proposed to reduce the experimental complexity.

By statistical analysis, several laws on relationships between similarity bias and distance, scale, translation and 3 axial rotations are found. We concluded that the rotations about two pattern axes mainly affect the geometrical bias of OCs. The fit limits of the bias under different rotation ranges are numerically approximated by the distribution of total sample space and extreme values.

The major contribution of our work is the discovery of a novel SLOC constraint which can be applicable for most practical situations. An application example was given as an instruction of how SLOC constraint can be applied in global optimization of

matching problem. The experimental results proved its effectiveness and efficiency.

The further work should focus on the several aspects described below.

- The precision of confirmatory experiments can be improved by designing better pattern, considering distortion of camera, increasing sample size, or utilizing real world pattern instead of MATLAB simulations.
- More application methods should be developed as SLOC constraint can be applied for many kinds of geometrical features.
- More hypothesis testing can be tried for other characteristics like distance or area size of region features as they may have the similar changing limits related to rotation angles as the orientation.

ACKNOWLEDGEMENT

This work is supported by the Key Project of Institute of Higher Education A-0282-21-207 from ZIME.

REFERENCES

- Alahi A, Ortiz R, Vandergheynst P (2012). FREAK: Fast retina keypoint. In CVPR:510–7.
- Balntas V, Lenc K, Vedaldi A, Mikolajczyk K (2017). HPatches: A benchmark and evaluation of handcrafted and learned local descriptors, In CVPR:3852–61.
- Bay H, Ess A, Tuytelaars T, Gool LV (2008). Speeded-up robust features (SURF). *Comput Vis Image Und* 110:346–59.
- Calonder M, Lepetit V, Strecha C (2010). BRIEF: Binary robust independent elementary features. In ECCV 6314:778–92.
- Fischler MA, Bolles RC (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 24: 381–95.
- Forsyth DA, Ponce J (2002). *Computer vision: a modern approach*. Prentice Hall Professional Technical Reference.
- Hamzah RA, Ibrahim H, Hassan AHA (2016). Stereo matching algorithm based on illumination control to improve the accuracy. *Image Anal Stereol* 35:39–52.
- Hu MK (1962). Visual pattern recognition by moment invariants. *IEEE T Inform Theory* 8:179–87.
- Lepetit V, Laguerre P, Fua P (2005). Randomized trees for real-time keypoint recognition. In CVPR 2: 775–81.
- Leutenegger S, Chli M, Siegwart RY (2011). BRISK: Binary robust invariant scalable keypoints, In ICCV 2548–55.
- Lieberknecht S, Benhimane S, Meier P, Navab N (2009). A dataset and evaluation methodology for template-based tracking algorithms. In ISMAR 145–51.
- Lowe DG (2004). Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* 60: 91–110.
- Mikolajczyk K, Schmid C (2005). A performance evaluation of local descriptors. *IEEE T Pattern Anal* 27:1615–30.
- Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, *et al.* (2005). A comparison of affine region detectors. *Int J Comput Vision* 65: 43–72.
- Valery N, Jesús A, Antonio A, Mossi JM, Soledad G (2007). Gradual transition detection for video partitioning using morphological operators. *Image Anal Stereol* 26:51–61.
- Ozuysal M, Fua P, Lepetit V (2007). Fast keypoint recognition in ten lines of code. In CVPR 1:1–8.
- Ozuysal M, Lepetit V, Fua P (2009). Pose estimation for category specific multiview object localization. In CVPR 778–85.
- Rublee E, Rabaud V, Konolige K, Bradski G (2011). ORB: An efficient alternative to SIFT or SURF. In ICCV 2564–71.
- Se S, Jasiobedzki P (2006). Photo-realistic 3D model reconstruction. In ICRA 3076–82.
- Tomažević D, Likar B, Pernuš F (2012). Multi-feature mutual information image registration. *Image Anal Stereol* 31:43–53.
- Xie T, Xie LJ, Song GH, Zheng Y (2013). Augmented reality registration from nature features based on planar color distribution. *J Zhejiang Univ (Eng Sci)* 47:2243–52.
- Zhou H, Yuan Y, Shi C (2009). Object tracking using SIFT features and mean shift. *Comput Vis Image Und* 113:345–52.