Direct 3D Pose Estimation of a Planar Target
Supplementary Material

Hung-Yu Tseng Po-Chen Wu
National Taiwan University
{hytseng,pcwu}@media.ee.ntu.edu.tw

Ming-Hsuan Yang
UC Merced
mhyang@ucmerced.edu

Shao-Yi Chien
National Taiwan University
sychien@ntu.edu.tw

1. Derivation Details of Bounded Steps in Table 1

For presentation clarity, in the following we use the notation c_a for $\cos(\theta_a)$ and s_a for $\sin(\theta_a)$ where $a = z_c$, x, and z_t. As discussed in (1) of the manuscript, the transformation between a reference point $\mathbf{x}_i = [x_i, y_i, 0]^\top$ and the corresponding image coordinate $\mathbf{u}_i = [u_i, v_i]^\top$ can be formulated as

$$\begin{bmatrix} hu_i \\ hv_i \\ h \end{bmatrix} = \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{R}|\mathbf{t}] \begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix}, \tag{1}$$

where

$$\mathbf{R} = \mathbf{R}_z(\theta_{z_c})\mathbf{R}_x(\theta_x)\mathbf{R}_z(\theta_{z_t}) = \begin{bmatrix} c_{z_c}c_{z_t} - c_x s_{z_c} s_{z_t} & -c_x c_{z_t} s_{z_c} - c_{z_c} s_{z_t} & s_x s_{z_c} \\ c_{z_t} s_{z_c} + c_x c_{z_c} s_{z_t} & c_x c_{z_c} c_{z_t} - s_{z_c} s_{z_t} & -s_x c_{z_c} \\ s_x s_{z_t} & s_x c_{z_t} & c_x \end{bmatrix}, \mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}. \tag{2}$$

It is shown in [2] that if

$$\forall \mathbf{x}_i \in I_t : d(T_{\mathbf{p}_1}(\mathbf{x}_i), T_{\mathbf{p}_2}(\mathbf{x}_i)) = O(\varepsilon), \tag{3}$$

then the following equation holds

$$|E_a(\mathbf{p}_1) - E_a(\mathbf{p}_2)| = O(\varepsilon\bar{\mathcal{V}}). \tag{4}$$

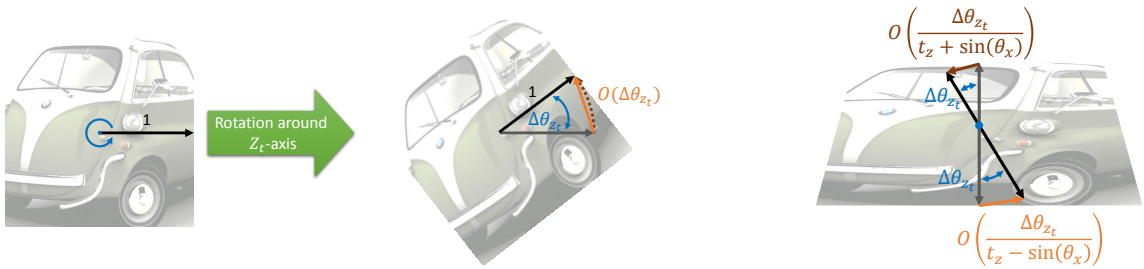
Our objective is to construct an ε-covering pose set $\mathcal{S}$. In this work, we construct $\mathcal{S}$ by first determining bounded steps for horizontal distance t_z and tilt angle θ_x. Then the bounded steps for the other dimensions θ_{z_c}, θ_{z_t}, t_x, and t_y can be determined based on t_z and θ_x.

Let $\theta_{z_t'} = \theta_{z_t} + \Delta\theta_{z_t}$, we obtain the following equation based on the current t_z and θ_x,

$$\begin{aligned} & d(T_{\mathbf{p}_{\theta_{z_t}}}(\mathbf{x}_i), T_{\mathbf{p}_{\theta_{z_t}+\Delta\theta_{z_t}}}(\mathbf{x}_i)) \\ &= \sqrt{f_x^2 \left[\frac{c_{z_t}x - s_{z_t}y}{s_x(s_{z_t}x + c_{z_t}y) + t_z} - \frac{c_{z_t'}x - s_{z_t'}y}{s_x(s_{z_t'}x + c_{z_t'}y) + t_z} \right]^2 + f_y^2 \left[\frac{c_x(s_{z_t}x + c_{z_t}y)}{s_x(s_{z_t}x + c_{z_t}y) + t_z} - \frac{c_x(s_{z_t'}x + c_{z_t'}y)}{s_x(s_{z_t'}x + c_{z_t'}y) + t_z} \right]^2} \\ &\leq \sqrt{f_x^2 \left[\frac{c_{z_t}x - s_{z_t}y}{s_x(s_{z_t}x + c_{z_t}y) + t_z} - \frac{c_{z_t'}x - s_{z_t'}y}{s_x(s_{z_t'}x + c_{z_t'}y) + t_z} \right]^2 + f_y^2 \left[\frac{s_{z_t}x + c_{z_t}y}{s_x(s_{z_t}x + c_{z_t}y) + t_z} - \frac{s_{z_t'}x + c_{z_t'}y}{s_x(s_{z_t'}x + c_{z_t'}y) + t_z} \right]^2} \\ &= O\left(\frac{\Delta\theta_{z_t}}{t_z + k\sin(\theta_x)} \right), \end{aligned} \tag{5}$$

where k denotes any constant in the range of $[-\sqrt{2}, \sqrt{2}]$. An illustrative example of (5) is shown in Figure 1. To make (5) satisfy the constraint in (3), we set the step size,

$$\Delta\theta_{z_t} = \Theta(\varepsilon(t_z + k\sin(\theta_x))), \tag{6}$$



(a) (b)

Figure 1. (a) 2D illustration of rotation around Z_t-axis. The linear distance (orange solid line) between points before and after applying rotation is bounded by the arc length (brown dotted line). (b) 3D illustration of rotation around Z_t-axis. The linear distance between points is a function of tilt angle θ_x.

where larger k means larger bounded steps for constructing $\mathcal{S}$. We set k to be 0 for $\Delta\theta_{z_t}$ in the proposed method.

As θ_{z_t} denotes 2D image rotation of the planar target, it does not affect the bounded steps for θ_{z_c}. Let $\theta_{z_c'} = \theta_{z_c} + \Delta\theta_{z_c}$, we obtain the following equation depending on the current t_z and θ_x, obtain the following equation depending on the current t_z and θ_x,

$$\begin{aligned} &d(T_{\mathbf{p}_{\theta_{z_c}}}(\mathbf{x}_i), T_{\mathbf{p}_{\theta_{z_c}+\Delta\theta_{z_c}}}(\mathbf{x}_i)) \\ &= \sqrt{f_x^2\left[\frac{c_{z_c}x - c_x s_{z_c} y}{s_x y + t_z} - \frac{c_{z_c'}x - c_x s_{z_c'} y}{s_x y + t_z}\right]^2 + f_y^2\left[\frac{s_{z_c}x + c_x c_{z_c} y}{s_x y + t_z} - \frac{s_{z_c'}x + c_x c_{z_c'} y}{s_x y + t_z}\right]^2} \\ &= O\left(\frac{\Delta\theta_{z_c}}{t_z + k\sin(\theta_x)}\right). \end{aligned} \tag{7}$$

We can realize (7) in a similar way to (5). To make (7) satisfy the constraint in (3), we set the step size,

$$\Delta\theta_{z_c} = \Theta(\varepsilon(t_z + k\sin(\theta_x))) = \Theta(\varepsilon(t_z)), \tag{8}$$

which k is setted to 0.

As the bounded steps for t_x and t_y are also affected by horizontal distance t_z and tilt angle θ_x only, the equations below can be derived,

$$\begin{aligned} &d(T_{\mathbf{p}_{t_x}(\mathbf{x}_i)}, T_{\mathbf{p}_{t_x+\Delta t_x}}(\mathbf{x}_i)) \\ &= \sqrt{f_x^2\left[\frac{x + t_x}{s_x y + t_z} - \frac{x + t_x + \Delta t_x}{s_x y + t_z}\right]^2 + f_y^2\left[\frac{y}{s_x y + t_z} - \frac{y}{s_x y + t_z}\right]^2} \\ &= O\left(\frac{\Delta t_x}{t_z + k\sin(\theta_x)}\right), \end{aligned} \tag{9}$$

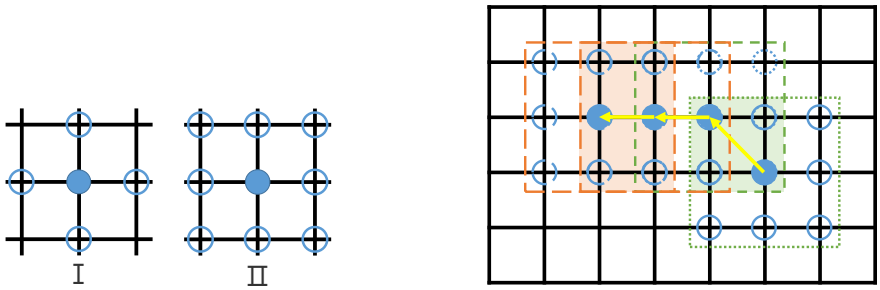
$$\begin{aligned} &d(T_{\mathbf{p}_{t_y}(\mathbf{x}_i)}, T_{\mathbf{p}_{t_y+\Delta t_y}}(\mathbf{x}_i)) \\ &= \sqrt{f_x^2\left[\frac{x}{s_x y + t_z} - \frac{x}{s_x y + t_z}\right]^2 + f_y^2\left[\frac{y + t_y}{s_x y + t_z} - \frac{y + t_y + \Delta t_y}{s_x y + t_z}\right]^2} \\ &= O\left(\frac{\Delta t_y}{t_z + k\sin(\theta_x)}\right), \end{aligned} \tag{10}$$

To make (9) and (10) satisfy the constraint in (3), we set these step sizes,

$$\Delta t_x = \Theta(\varepsilon(t_z + k\sin(\theta_x))) = \Theta(\varepsilon(t_z + \sqrt{2}\sin(\theta_x))), \tag{11}$$

$$\Delta t_y = \Theta(\varepsilon(t_z + k\sin(\theta_x))) = \Theta(\varepsilon(t_z + \sqrt{2}sin(\theta_x))), \tag{12}$$

as k is setted to $\sqrt{2}$ for pratical consideration.



(a) (b)

Figure 2. (a) Two types of checking patterns in 2D view. (b) 2D itinerary illustration of gradient descent search. The next step is centered on either corner point or edge point. In the first case (centered on corner point), 44% (namely, $\frac{4}{9}$) checking points have already been visited; in the second case (centered on edge point), 67% (namely, $\frac{6}{9}$) checking points have already been visited.

2. Different Types of Checking Patterns

We design two types of checking patterns in the gradient descent search scheme [4] for the 6D parameter space. The 2D view of these candidate checking patterns are shown in Figure 2(a). Type I is formed around the center point, and the 12 neighbors are ε-away from the center separately in the 6D pose space, and is used in the proposed algorithm. Type II is composed of 729 (i.e., 3^6) checking points, including three different values (i.e., $[+\Delta(\varepsilon), 0, -\Delta(\varepsilon)]$, where $\Delta(\varepsilon)$ represents the bounded step) per dimension. Note that after the first iteration, many of the checking points in Type II would have been checked in the previous step, as the 2D illustration shown in Figure 2(b). The percentage of visited checking points is distributed from 8.8% (corner) to 66.7% (edge). In spite of this property, the number of unchecked points in each iteration is still too large, therefore the refinement scheme cannot converge in a short time. In our experiments, it will take more than one hour to converge with Type II during the refinement process. In this work ,we use Type I in the despite the probability of being trapped in a local minimum is high. We address this issue by using the efficient and effective coarse-to-fine scheme.

3. Experimental Settings

3.1. Synthetic Dataset

The template images in synthetic image experiment are acquired from the dataset by Lieberknecht *et al*. [3]. While this dataset also includes some videos with different motion patterns for evaluating pose estimation and tracking algorithms, the ground truth is not available. The test images under different kinds of degradation are generated using MATLAB built in functions.

3.2. Benchmark Dataset

As the benchmark dataset by Gauglitz *et al*. [1] only provides homography parameters of the videos, we have to calculate the 6D poses from these homography parameters by ourselves. This process is conducted as follows:

1. Straighten the camera images with calibration coefficients.
2. Modify the texture margin from 1.5 to 0.75 (the original texture margin is incorrect).
3. Explore the 4 correspondences of template corners with provides homography parameters.
4. Compute the candidate poses with these 4 correspondences using OPnP [5].
5. Select the correct pose from all candidate poses.

Similar to [1], these pose parameters are used as the ground truth for performance evaluation. In addition, some homography parameters of some video sequences (i.e., “br-m1”, “pa-pn”, etc.) are recorded in the wrong way, we manually correct all of them.

4. Additional Experimental Results

4.1. Synthetic Images

Figure 3, 5, 6, 7, and 8 show the pose estimation results using different approaches on synthetic images under varying conditions. The whole test conditions are: normal case, Gaussian blur with standard deviation $\{1,2,3,4,5\}$, JPEG compression

Table 1. Experimental results under varying conditions. The SIFT-based (S), ASIFT-based (A), and the proposed direct (D) methods are evaluated under different conditions (B: Gaussian blur; J: JPEG compression; I: intensity change; T: tile angle change) with degradation level from 1 to 5 .

		Bump Sign			Stop Sign			Lucent			MacMini Board			Isetta			Philadelphia			Grass			Wall		
		$E_R(°)$	$E_t(\%)$	SR(%)	$E_R(°)$	$E_t(\%)$	SR(%)	$E_R(°)$	$E_t(\%)$	SR(%)	$E_R(°)$	$E_t(\%)$	SR(%)	$E_R(°)$	$E_t(\%)$	SR(%)	$E_R(°)$	$E_t(\%)$	SR(%)	$E_R(°)$	$E_t(\%)$	SR(%)	$E_R(°)$	$E_t(\%)$	SR(%)
B1	S	108	49.8	6.00	103	60.2	10.0	106	99.8	16.0	68.4	69.9	34.0	53.8	24.5	50.0	19.1	9.64	82.0	113	280	4.00	24.3	30.3	84.0
	A	50.9	19.5	30.0	4.29	0.79	98.0	5.68	0.93	94.0	4.70	1.17	96.0	5.44	0.88	92.0	2.07	0.40	98.0	66.2	15.5	46.0	1.41	0.37	98.0
	D	4.20	0.80	96.0	2.31	1.04	98.0	2.13	0.92	100.0	9.86	3.43	90.0	2.00	0.79	98.0	3.77	0.90	98.0	4.97	1.68	96.0	5.59	1.17	92.0
B2	S	117	56.0	0.00	120	60.6	2.00	132	114	0.00	127	181	2.00	94.5	53.9	16.0	87.7	46.3	22.0	123	285	0.00	100	351	16.0
	A	64.1	24.7	24.0	4.53	0.94	96.0	3.12	0.99	94.0	6.32	1.18	90.0	3.93	0.74	96.0	1.94	0.50	100	77.3	26.1	30.0	13.0	3.13	90.0
	D	2.71	0.93	98.0	1.73	1.05	100	2.19	1.05	98.0	9.20	3.38	84.0	3.09	0.84	96.0	3.66	1.22	96.0	4.46	1.78	94.0	4.50	1.55	96.0
B3	S	112	64.1	0.00	118	72.8	0.00	124	180	0.00	129	669	0.00	129	71.3	2.00	126	89.9	2.00	126	669	0.00	119	968	4.00
	A	75.4	29.5	18.0	25.4	6.26	70.0	29.1	9.24	76.0	44.9	20.0	58.0	23.8	11.2	76.0	14.5	9.98	80.0	103	35.1	6.00	40.1	12.0	64.0
	D	4.42	1.24	94.0	3.51	1.18	96.0	5.71	2.99	92.0	16.9	5.06	76.0	2.95	1.25	98.0	3.25	1.07	98.0	6.48	2.41	92.0	6.31	1.77	96.0
B4	S	120	53.4	0.00	125	76.0	0.00	127	398	0.00	130	459	0.00	123	62.6	0.00	128	163	0.00	115	693	0.00	129	715	0.00
	A	94.2	36.1	10.0	46.8	15.1	60.0	45.9	18.5	46.0	90.8	29.7	20.0	47.8	14.9	56.0	59.2	20.2	46.0	110	35.1	0.00	82.4	27.1	26.0
	D	3.82	1.37	98.0	5.42	1.44	96.0	4.62	2.14	96.0	9.99	4.08	84.0	3.10	1.55	96.0	7.99	1.42	92.0	11.0	4.17	86.0	11.3	2.81	82.0
B5	S	127	72.4	0.00	124	87.5	0.00	129	764	0.00	129	894	0.00	128	58.1	0.00	129	131	0.00	124	821	0.00	123	1316	0.00
	A	93.1	38.1	0.00	88.9	27.6	28.0	72.0	26.9	28.0	88.7	27.3	20.0	90.5	26.2	18.0	60.2	21.3	40.0	105	34.0	0.00	96.3	33.2	12.0
	D	2.96	1.48	100	3.43	1.73	96.0	4.50	2.47	96.0	16.5	2.06	80.0	2.68	2.08	100	12.3	2.02	86.0	17.8	5.96	76.0	25.1	6.66	68.0
J1	S	112	60.5	2.00	54.3	29.2	52.0	44.3	36.7	66.0	31.6	50.8	74.0	26.9	12.8	78.0	7.76	4.55	94.0	84.3	133	34.0	29.6	21.9	80.0
	A	56.7	22.6	32.0	3.90	0.74	98.0	1.58	0.40	98.0	3.18	1.32	96.0	3.40	0.52	96.0	0.88	0.29	100	70.2	22.4	36.0	3.97	0.71	98.0
	D	1.58	0.81	98.0	2.14	1.00	100	3.54	1.63	94.0	10.83	1.50	92	1.17	0.76	100	4.00	1.00	94.0	5.27	1.34	92.0	4.68	1.43	94.0
J2	S	110	61.1	4.00	57.9	26.7	46.0	32.8	25.1	76.0	22.3	9.73	82.0	40.3	15.8	64.0	15.2	15.6	90.0	75.9	109	36.0	21.3	11.7	86.0
	A	70.0	27.9	28.0	8.14	0.81	94.0	1.16	0.36	100	2.96	1.59	96.0	5.50	1.11	94.0	0.97	0.42	100	49.1	18.4	52.0	1.29	0.30	100
	D	3.65	0.80	94.0	1.95	0.87	100	3.72	1.00	96.0	5.48	1.98	96.0	1.55	0.91	100	1.99	1.11	100	2.37	2.35	94.0	4.91	1.21	94.0
J3	S	109	52.4	6.00	57.7	27.6	52.0	50.8	28.0	62.0	25.9	11.3	80.0	32.8	13.4	74.0	6.45	3.69	96.0	95.5	45.6	26.0	18.1	47.5	86.0
	A	46.9	18.8	32.0	6.02	0.82	94.0	2.08	0.39	98.0	6.02	0.46	96.0	3.57	0.53	96.0	1.78	0.33	98.0	56.7	21.2	48.0	1.31	0.35	100
	D	2.39	0.82	100	1.78	0.83	100	1.82	1.02	100	6.91	1.83	94.0	1.89	0.76	100	3.08	1.06	98.0	3.76	1.67	94.0	3.83	0.98	98.0
J4	S	119	60.9	4.00	47.5	19.8	56.0	39.9	21.1	66.0	30.2	29.7	76.0	38.8	16.2	70.0	18.2	8.91	86.0	91.6	115	28.0	25.2	55.1	84.0
	A	68.1	24.0	24.0	3.09	0.64	96.0	2.30	0.42	98.0	4.38	0.50	96.0	5.70	0.95	94.0	1.06	0.28	98.0	49.4	14.1	60.0	4.71	0.67	96.0
	D	2.39	0.90	98.0	1.93	0.83	100	2.74	1.04	100	15.4	4.39	82.0	1.20	0.80	100	1.33	0.73	100	4.93	7.76	90.0	7.44	1.41	94.0
J5	S	117	57.8	2.00	43.9	27.4	60.0	35.7	16.9	68.0	24.8	24.9	80.0	31.4	19.6	74.0	30.0	6.37	88.0	101.2	101	20.0	10.2	4.41	92.0
	A	45.4	21.3	42.0	6.61	7.33	94.0	1.19	0.38	100	4.22	1.37	98.0	3.57	0.63	94.0	0.97	0.29	100	67.0	24.7	36.0	1.74	0.37	100
	D	2.75	1.00	98.0	2.29	0.93	98.0	2.96	1.39	96.0	2.92	2.60	94.0	2.30	0.87	98.0	1.55	0.86	100	7.06	1.51	90.0	5.30	2.89	92.0
I1	S	104	48.1	4.00	45.0	22.6	58.0	40.6	85.1	64.0	29.8	59.1	74.0	24.8	12.4	76.0	16.2	6.72	88.0	100.7	170	20.0	11.6	23.2	90.0
	A	53.4	17.5	32.0	6.64	0.84	92.0	2.48	0.42	98.0	7.60	1.48	92.0	3.06	0.45	96.0	1.41	0.36	100	73.0	22.4	38.0	0.81	0.26	100.0
	D	2.64	0.85	98.0	3.14	1.05	98.0	5.49	1.55	94.0	5.06	3.55	96.0	2.29	0.86	98.0	3.12	1.09	98.0	2.26	1.08	98.0	1.56	0.81	98.0
I2	S	107	53.4	12.0	45.0	23.9	58.0	24.8	34.7	80.0	23.0	20.6	80.0	24.4	12.3	80.0	13.5	7.01	90.0	76.9	96.9	36.0	26.4	34.7	78.0
	A	61.8	27.6	26.0	11.5	2.59	94.0	1.17	0.43	100	3.82	0.52	94.0	5.59	0.71	94.0	0.96	0.32	100	83.9	28.1	28.0	1.87	0.43	98.0
	D	1.15	0.60	100	2.07	0.81	98.0	1.87	0.69	100	4.66	1.77	96.0	1.64	0.87	100	2.07	1.05	100	5.06	1.90	90.0	1.64	0.92	100
I3	S	116	64.7	2.00	53.0	28.7	50.0	29.8	25.9	76.0	20.4	31.3	84.0	24.3	8.00	82.0	20.7	7.27	84.0	82.0	80.8	36.0	4.92	9.08	96.0
	A	58.8	20.6	18.0	3.64	0.80	98.0	2.92	0.40	96.0	1.77	0.47	98.0	1.76	0.47	100	1.83	0.52	98.0	64.3	32.6	34.0	1.26	0.32	100
	D	1.66	0.82	100	2.55	0.93	100	7.52	2.74	94.0	2.11	1.30	98.0	2.59	0.74	98.0	3.30	1.32	96.0	5.35	3.55	90.0	3.52	1.20	96.0
I4	S	114	59.3	6.00	35.4	9.80	68.0	38.6	29.9	66.0	13.2	15.4	90.0	27.6	13.3	74.0	10.9	3.56	92.0	67.5	67.7	48.0	12.0	22.8	88.0
	A	73.1	32.5	18.0	4.82	0.97	94.0	4.42	1.00	96.0	5.70	1.34	96.0	6.09	0.61	94.0	2.38	0.39	98.0	83.0	24.9	28.0	8.75	2.11	94.0
	D	2.40	0.68	98.0	2.56	0.70	98.0	3.91	0.86	94.0	9.83	1.42	88.0	2.44	0.82	98.0	2.42	1.05	98.0	7.27	3.44	84.0	5.14	4.24	96.0
I5	S	117	61.7	2.00	59.7	31.6	46.0	60.8	184	48.0	34.7	32.4	74.0	31.5	14.0	76.0	14.0	10.4	90.0	41.1	68.1	32.0	15.9	19.1	88.0
	A	80.3	37.3	10.0	7.64	1.06	94.0	18.7	4.69	84.0	6.76	1.06	96.0	4.91	0.67	94.0	4.31	0.44	96.0	97.3	33.2	18.0	0.62	1.17	96.0
	D	3.93	1.01	96.0	2.88	0.77	98.0	6.12	7.80	94.0	5.59	5.52	94.0	1.52	1.04	98.0	2.12	1.29	98.0	9.72	7.44	84.0	4.29	1.44	98.0
T1	S	105	50.7	12.0	22.7	9.43	80.0	4.97	6.58	96.0	0.44	0.17	100	6.37	2.55	96.0	1.61	0.22	100	80.2	69.1	30.0	0.88	0.20	100
	A	41.4	15.7	34.0	9.27	0.63	86.0	2.93	0.34	98.0	4.67	0.42	96.0	5.19	0.46	92.0	3.91	0.37	96.0	56.0	24.3	42.0	2.91	0.30	100
	D	5.59	0.80	94.0	5.33	1.22	98.0	3.64	0.72	96.0	3.98	0.88	98.0	3.42	0.71	98.0	6.30	1.40	94.0	9.22	1.45	88.0	11.6	1.50	88.0
T2	S	99.6	56.1	8.00	43.8	24.0	62.0	5.54	5.56	94.0	0.35	0.18	100	8.66	2.12	92.0	0.35	0.20	100	44.2	21.5	62.0	0.34	0.20	100
	A	43.6	14.9	36.0	4.74	0.66	94.0	2.46	0.33	96.0	2.60	0.42	98.0	3.86	0.51	94.0	0.77	0.27	100	45.4	14.3	60.0	0.97	0.27	100
	D	2.72	0.94	100	2.61	0.98	100	2.11	0.82	98	9.18	1.28	88.0	1.77	0.89	100	2.24	0.89	100	6.25	1.06	92.0	16.1	1.80	82.0
T3	S	108	64.8	6.00	57.1	24.7	56.0	15.3	11.6	86.0	0.24	0.20	100	16.2	8.09	88.0	0.29	0.25	100	66.9	49.7	44.0	2.07	83.5	98.0
	A	46.0	19.4	44.0	2.57	0.67	98.0	0.53	0.29	100	4.12	0.42	96.0	2.20	0.38	98.0	0.58	0.31	100	34.8	11.1	66.0	0.64	0.34	100
	D	1.27	0.72	100	1.16	1.12	100	0.88	0.77	100	7.87	1.34	94.0	1.06	0.78	100	1.59	0.99	100	0.76	0.90	100	1.51	0.93	98.0
T4	S	118	88.5	0.00	94.3	52.5	18.0	76.3	207	40.0	45.5	34.8	66.0	51.2	22.8	60.0	7.11	3.76	92.0	114	120	18.0	15.0	14.5	86.0
	A	58.2	23.2	40.0	4.91	0.81	96.0	0.60	0.43	100	4.74	0.47	96.0	0.64	0.51	100	0.53	0.42	100	54.5	16.3	50.0	0.59	0.35	100
	D	0.97	0.73	100	0.70	0.79	100	0.69	1.02	100	1.34	1.28	98.0	0.55	0.73	100	1.03	0.96	100	0.80	1.03	100	1.49	1.15	98.0
T5	S	116	51.7	0.00	124	57.2	0.00	130	97.0	0.00	128	145	0.00	115	58.3	4.00	90.9	33.0	34.0	129	116	0.00	97.6	127	18.0
	A	84.5	39.5	12.0	6.54	1.71	96.0	0.61	0.65	100	14.0	3.92	90.0	3.73	0.84	98.0	0.45	0.48	100	87.5	23.0	28.0	2.59	0.89	98.0
	D	0.41	0.73	100	0.42	0.79	100	0.92	0.85	100.0	16.2	8.83	86.0	0.41	0.71	100	0.64	0.66	100	1.24	2.97	94.0	1.14	1.47	98.0

with quality parameter {90,80,70,60,50}, intensity change with scalar parameter {0.9,0.8,0.7,0.6,0.5}, and tile angle in the range {[0°15°), [15°30°), [30°45°), 45°60°), and [60°75°)}. The complete comparison results under varying conditions using two feature-based methods and the proposed direct algorithm are shown in Table 1.

4.2. Real Images

Figure 9, 10, 11, and 12 show the pose estimation results on the benchmark dataset [1] by different approaches. Other results on this benchmark can also be found in the supplementary video.

As the template images provided by these datasets [3, 1] are designed for evaluating feature-based approaches, they are not textureless. Figure 4 shows some pose estimation results with textureless template images using the proposed direct method. Here we do not show the results using feature-based approaches because both of them cannot estimate pose correctly due to lack of sufficient feature correspondences. These results demonstrate that our method can perform pose estimation robustly not only on targets with rich texture, but on textureless targets. Other results can also be found in the supplementary video.

References

[1] S. Gauglitz, T. Höllerer, and M. Turk. Evaluation of interest point detectors and feature descriptors for visual tracking. *IJCV*, 94(3):335–360, 2011. 3, 4

Figure 3. Pose estimation results on synthetic images by SIFT-based approach (gray boxes), ASIFT-based approach (magenta boxes), and proposed direct method (cyan boxes) under normal condition.

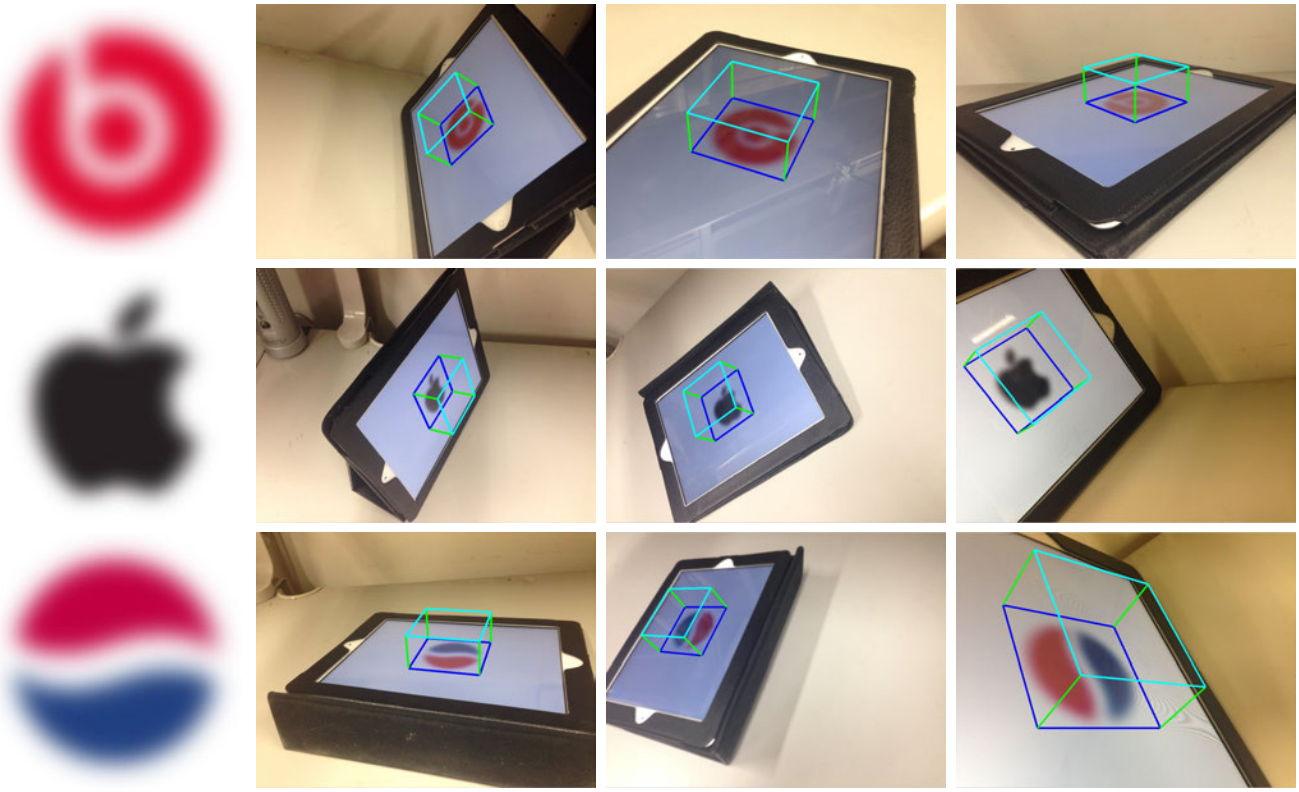
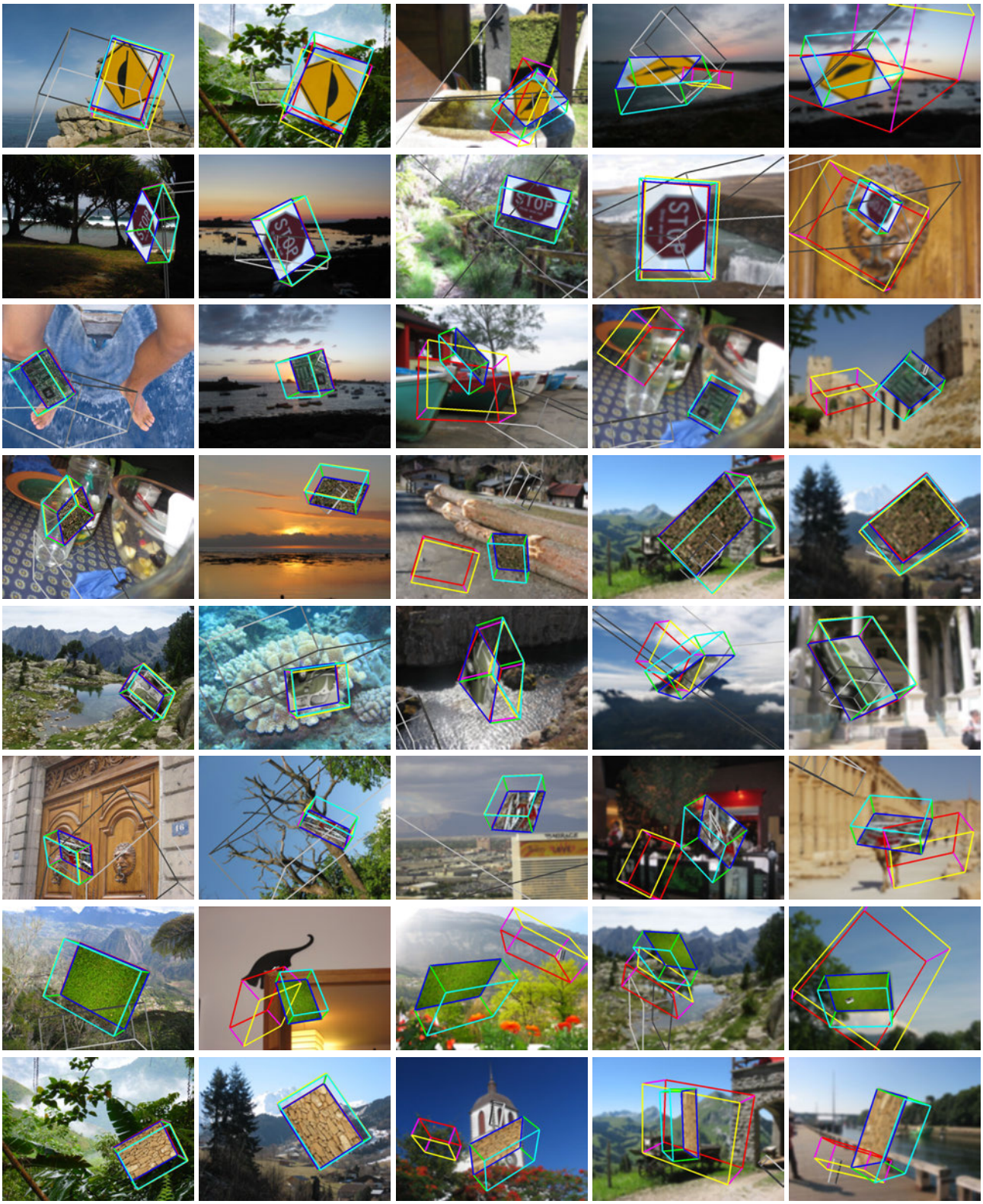


Figure 4. Pose estimation results on textureless images by proposed direct method. First column: pose estimation target. Other columns: pose estimation results with rendered box.

[2] S. Korman, D. Reichman, G. Tsur, and S. Avidan. Fast-match: Fast affine template matching. In *CVPR*, 2013. 1

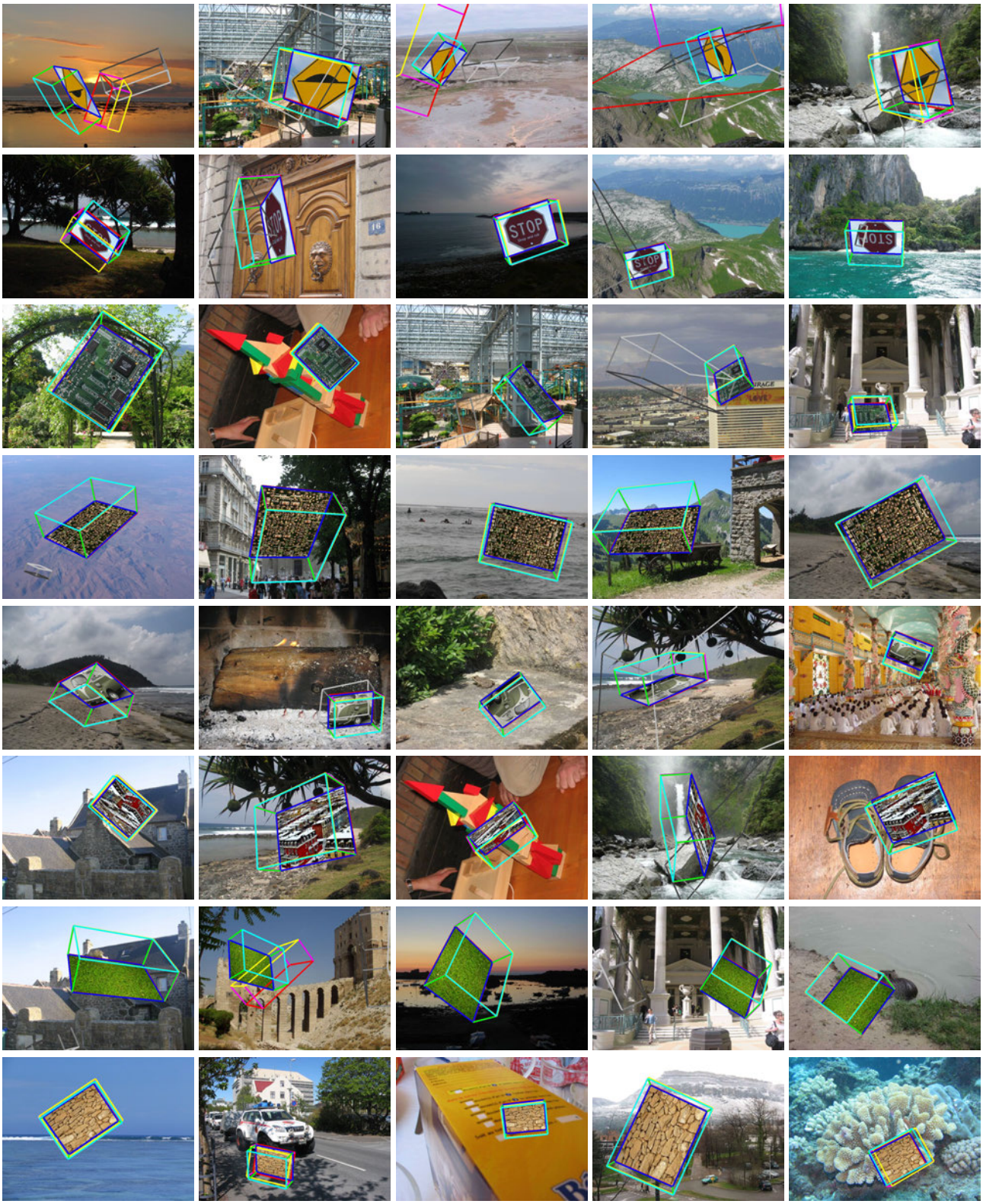
[3] S. Lieberknecht, S. Benhimane, P. Meier, and N. Navab. A dataset and evaluation methodology for template-based tracking algorithms. In *ISMAR*, 2009. 3, 4

[4] L.-K. Liu and E. Feig. A block-based gradient descent search algorithm for block motion estimation in video coding. *IEEE TCSVT*, 6(4):419–422, 1996. 3



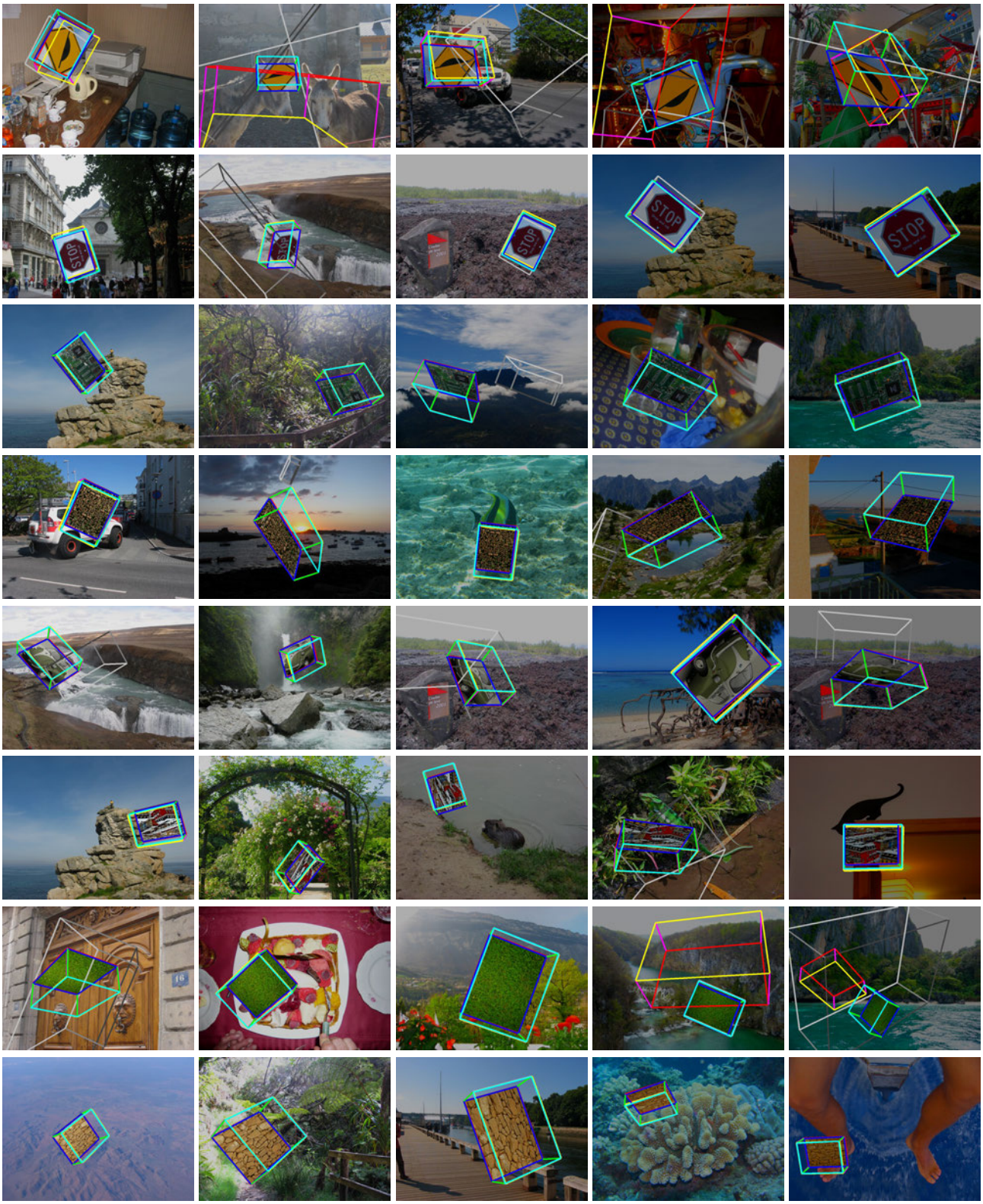
Kernel Size 1.0 | Kernel Size 2.0 | Kernel Size 3.0 | Kernel Size 4.0 | Kernel Size 5.0

Figure 5. Pose estimation results on synthetic images by SIFT-based approach (gray boxes), ASIFT-based approach (magenta boxes), and proposed direct method (cyan boxes) under Gaussion blur condition.



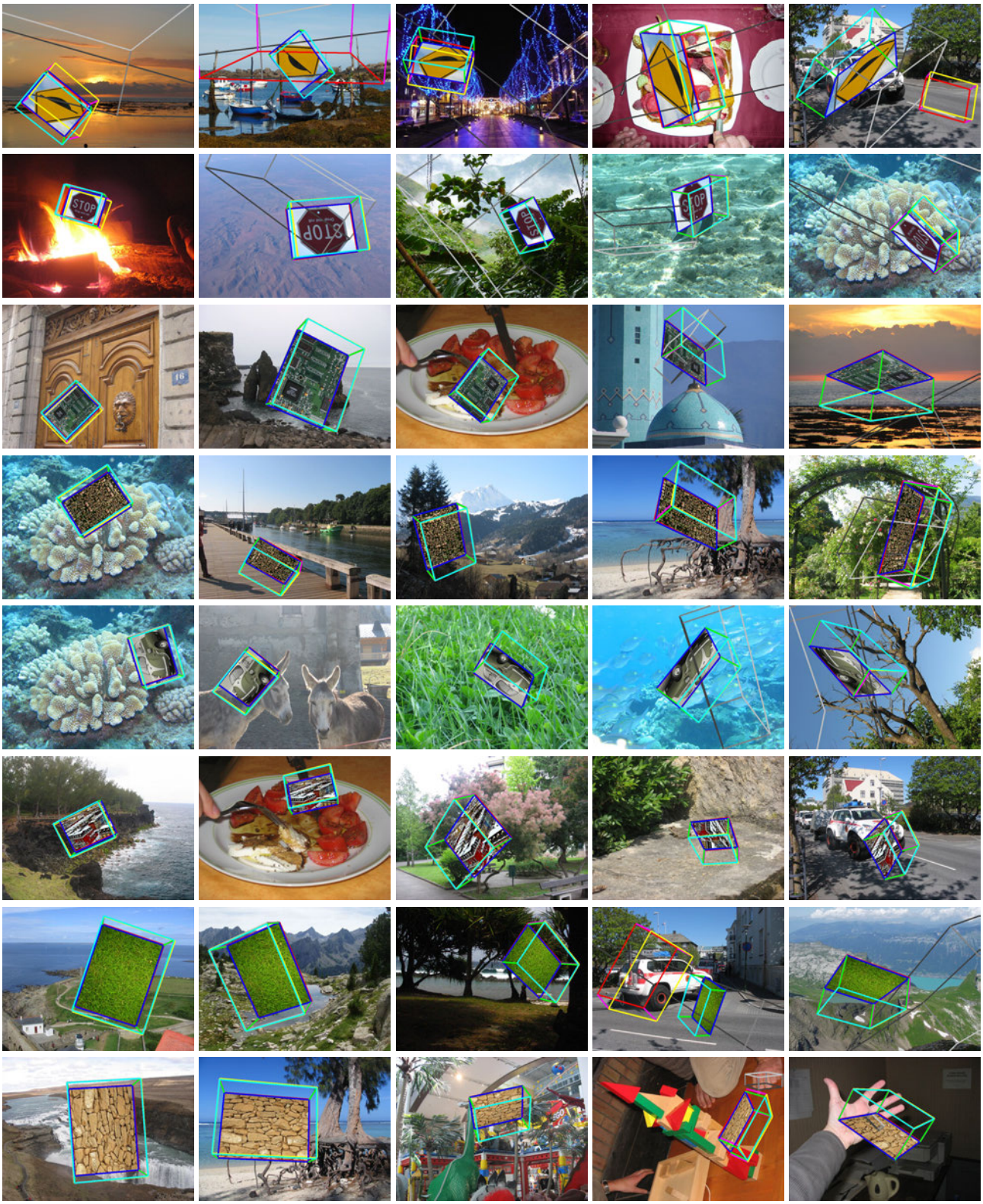
Quality Parameter 90 | Quality Parameter 80 | Quality Parameter 70 | Quality Parameter 60 | Quality Parameter 50

Figure 6. Pose estimation results on synthetic images by SIFT-based approach (gray boxes), ASIFT-based approach (magenta boxes), and proposed direct method (cyan boxes) under JPEG compression.



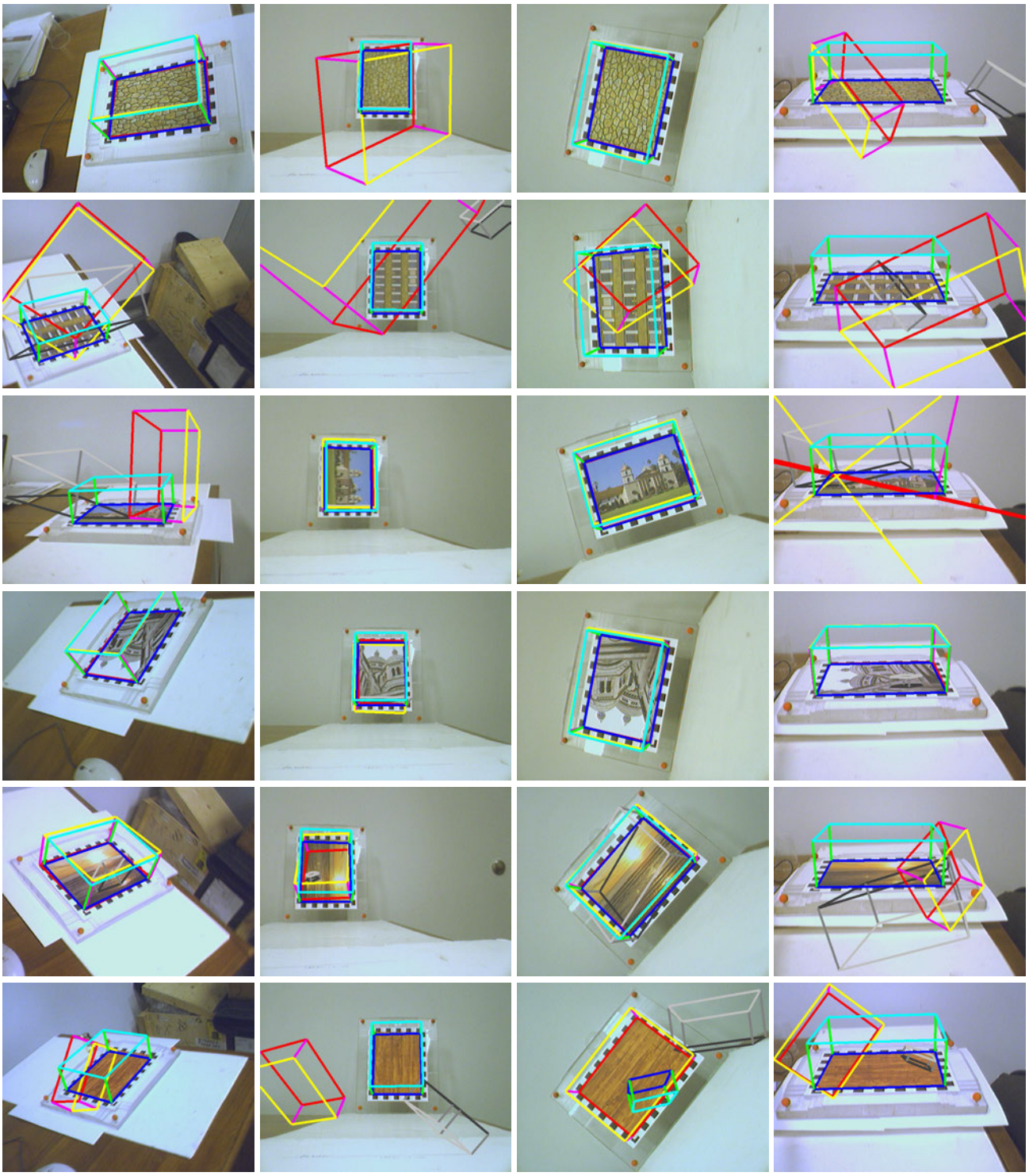
Intensity Scale 0.9 Intensity Scale 0.8 Intensity Scale 0.7 Intensity Scale 0.6 Intensity Scale 0.5

Figure 7. Pose estimation results on synthetic images by SIFT-based approach (gray boxes), ASIFT-based approach (magenta boxes), and proposed direct method (cyan boxes) under intensity change condition.



Tile Angle [0° 15°) Tile Angle [15° 30°) Tile Angle [30° 45°) Tile Angle [45° 60°) Tile Angle [60° 75°)

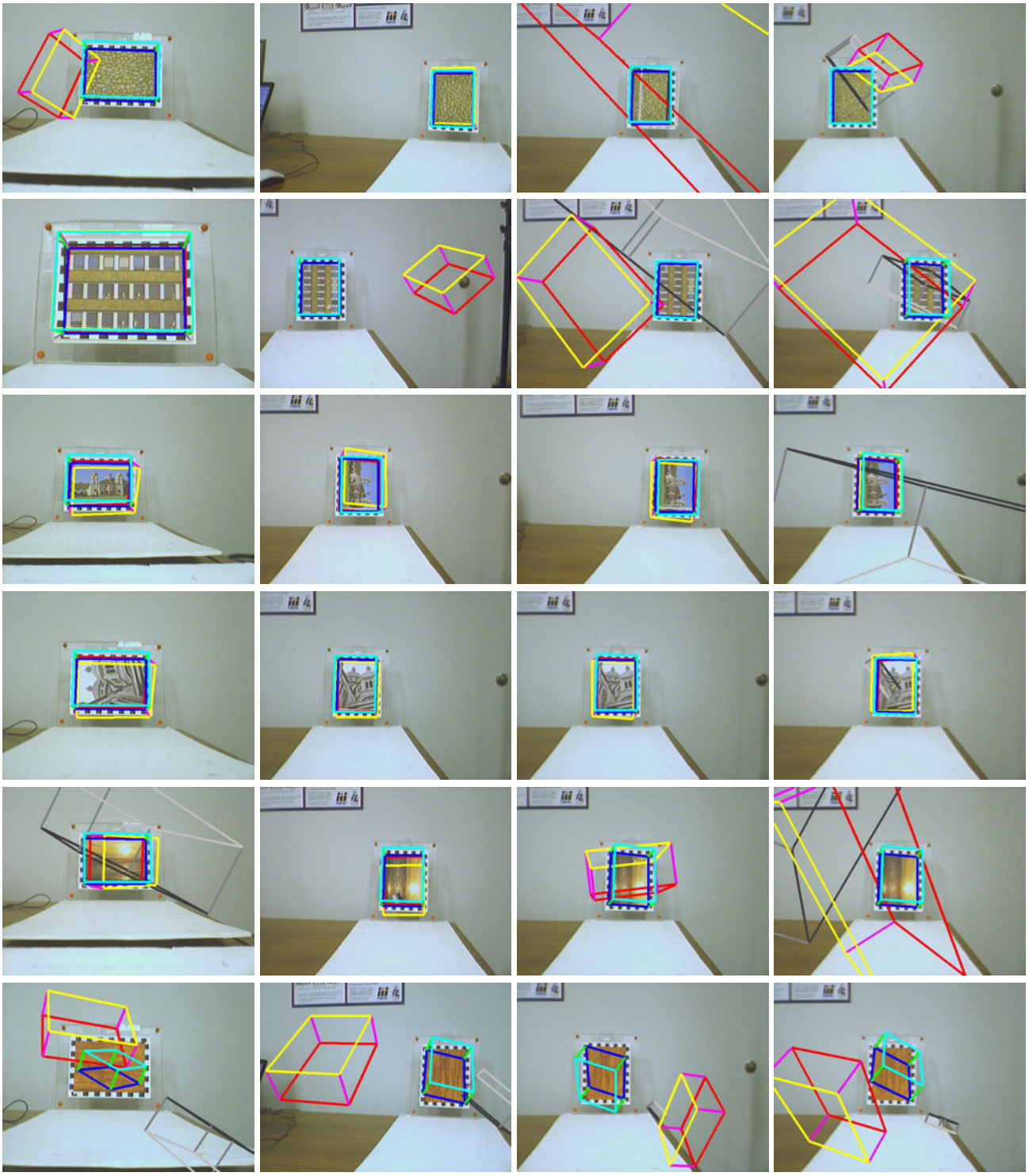
Figure 8. Pose estimation results on synthetic images by SIFT-based approach (gray boxes), ASIFT-based approach (magenta boxes), and proposed direct method (cyan boxes) under different tilt angles.



Unconstrained Panning Rotation Perspective Distortion

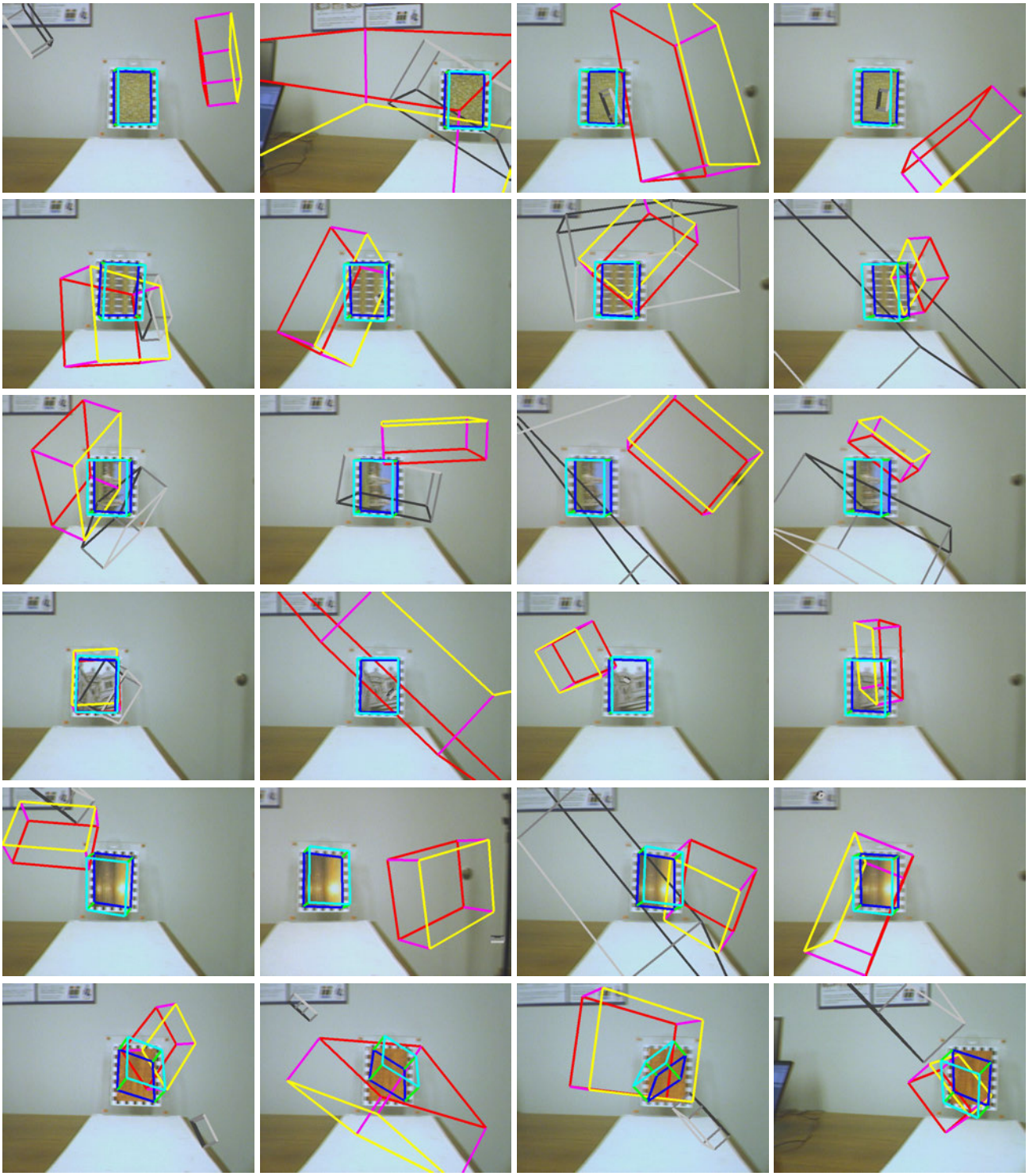
Figure 9. Pose estimation results on benchmark dataset by SIFT-based approach (gray boxes), ASIFT-based approach (magenta boxes), and proposed direct method (cyan boxes).

[5] Y. Zheng, Y. Kuang, S. Sugimoto, K. Astrom, and M. Okutomi. Revisiting the PnP Problem: A Fast, General and Optimal Solution. In *ICCV*, 2013. 3



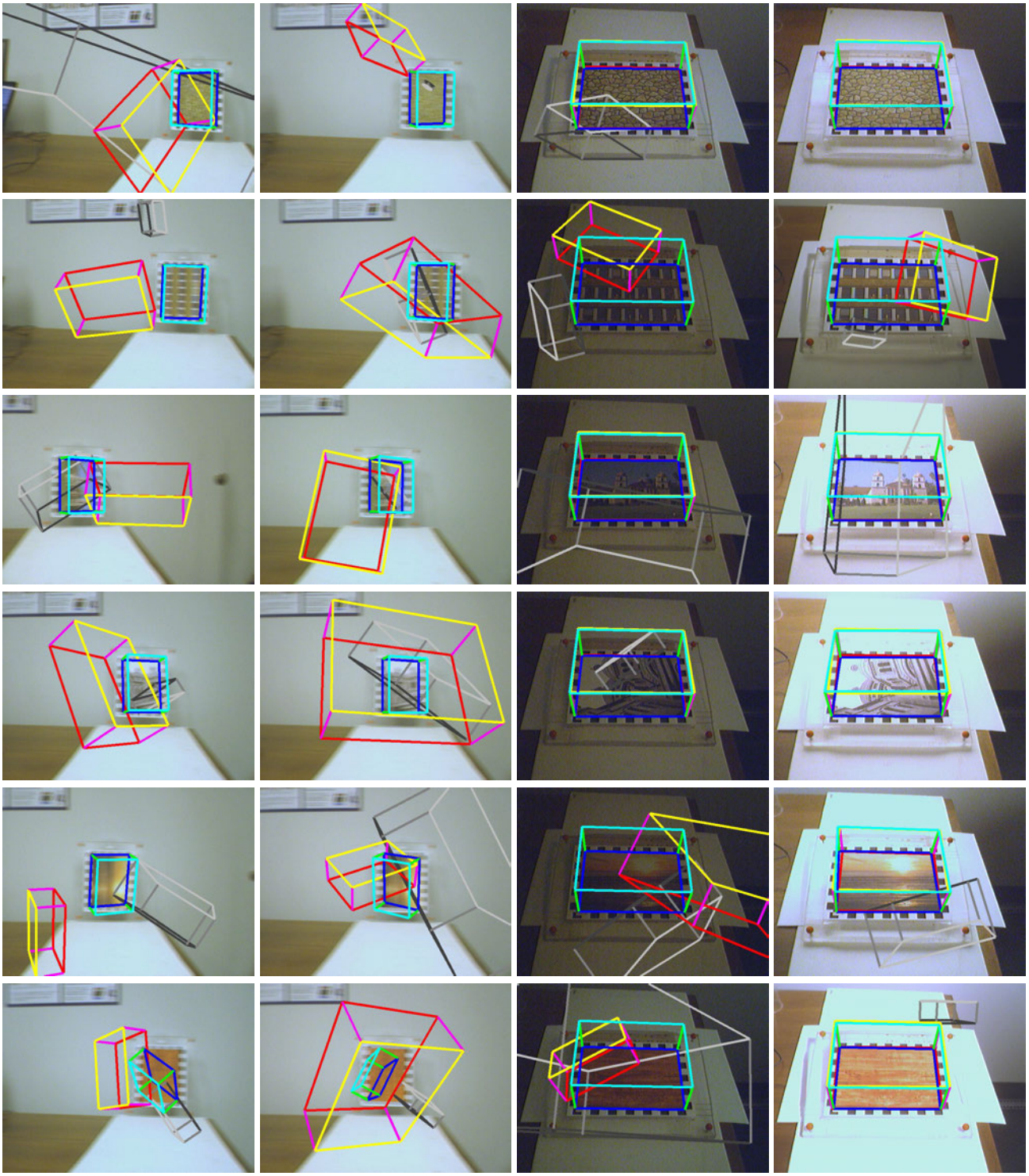
Zoom | Motion Blur (Level 1) | Motion Blur (Level 2) | Motion Blur (Level 3)

Figure 10. Pose estimation results on benchmark dataset by SIFT-based approach (gray boxes), ASIFT-based approach (magenta boxes), and proposed direct method (cyan boxes).



Motion Blur (Level 4) Motion Blur (Level 5) Motion Blur (Level 6) Motion Blur (Level 7)

Figure 11. Pose estimation results on benchmark dataset by SIFT-based approach (gray boxes), ASIFT-based approach (magenta boxes), and proposed direct method (cyan boxes).



Motion Blur (Level 8) | Motion Blur (Level 9) | Static Lighting | Dynamic Lighting

Figure 12. Pose estimation results on benchmark dataset by SIFT-based approach (gray boxes), ASIFT-based approach (magenta boxes), and proposed direct method (cyan boxes).