VISUAL TRACKING USING BLIND SOURCE SEPARATION FOR MIXED IMAGES

*Hsiao-Tzu Chen and Chih-Wei Tang**

Department of Communication Engineering, National Central University, Jhongli, Taiwan
{tzuchen248@gmail.com, cwtang@ce.ncu.edu.tw}

ABSTRACT

Mixed images cannot be avoided in visual tracking since the transmitted scene may be captured with specular reflections. Since few previous methods tackle this important problem, this paper proposes a novel visual tracking method using Blind Source Separation (BSS) for mixed images. Based on the framework of particle filter with compensated motion model at the prediction stage for mobile cameras, this paper improves its correction stage by weighting particles using color histograms on the mixed image and intrinsic illumination image, based on the trichromatic and opponent-process theories, respectively. Moreover, the weighting of each particle is optimized using Maximum Likelihood (ML). Experimental results show that the proposed scheme effectively improves the tracking accuracy on mixed images.

***Index Terms*—**Visual tracking, reflection, blind source separation, particle filter, correction stage

1. INTRODUCTION

Previous methods of visual tracking tackle mostly problems including partial occlusion, illumination variations, complex trajectory, camera motion, etc. To the best of our knowledge, few of them focus on reflection problem. In fact, for mixed images caused by specular reflections, inaccurate tracking results are easily raised due to the significant change of appearances of targets and/or the background.

To improve the accuracy of a probabilistic based tracker in a dynamic or cluttered environment, a richer representation of the target may help the correction stage. For example, Wu et al [1] propose efficient and robust co-inference tracking using two modalities of the target. Jin et al [2] propose an object model, mixing non-parametric contour and edge models. Tracking can be also casting as finding a sparse approximation in a set of the target template and trivial templates [3]. Such method is proved to be quite robust against various conditions such as occlusion.

Blind source separation estimates source signals with an unknown mixing matrix from a set of mixed signals. For a mixed image including an intrinsic component and the specular reflection, separation methods use the linear mixing image formulation and assume that source signals are independent. Some of the reflection separation approaches are based on Independent Component Analysis (ICA) (e.g., [4]). They need at least two static mixed images with diverse conditions. Given two initial mixtures, it is proposed to separate two sources by minimizing their structural correlations [5]. Inspired by Sarel et al [5], the multiple generalized normalized gray-scale correlation assists iterative estimation of multiple sources [6]. Estimation can be under the sparse prior over derivative filters on natural images [7][8]. Separation from a single image can be achieved using manually marked small amount of edges [8] or minimizing the total amount of edges and corners [9]. Gai et al [10] further consider the diversities of layer motions. Instead of frame based reflection detection, the method in [11] uses tracking to detect reflectance regions in video frames. In addition, it proposes the concept of applying separation before object tracking in mixed images.

Different from previous tracking methods and applications, this paper proposes a novel tracking method for mixed images with two layers, including the target and the reflection layers. Based on the framework of particle filter with motion compensated motion model at the prediction stage for mobile cameras [12], this paper improves the correction stage by weighting each particle using ML, where both the color histograms on the mixed and the intrinsic illumination images are observed. The intrinsic images are derived by Weiss' separation method [7], not iterative and more suitable for real-time tracking.

This paper is organized as follows. Section 2 and Section 3 review the compensated motion model of visual tracking for mobile cameras [12] and the Weiss' reflection separation [7], respectively. Section 4 proposes a novel method for visual tracking in mixed images and using ML to optimize the weight of each particle. Section 5 gives experimental results and Section 6 concludes this paper.

2. VISUAL TRACKING USING COMPENSATED MOTION MODEL FOR MOBILE CAMERAS

The particle filter (PF) implements the Bayesian filter recursively using the Monte Carlo method [13]. Bayesian tracking consists of prediction and correction to estimate the state over the posterior pdf. Prediction obtains the prior pdf of the state, $\mathbf{x}_t$, at time t by

$$p(\mathbf{x}_t \mid \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} \mid \mathbf{z}_{1:t-1}) d\mathbf{x}_{t-1} \quad (1)$$

where $\mathbf{z}_{1:t-1} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{t-1}\}$ is the set of all observations up to time t-1. The correction stage updates the posterior density $p(\mathbf{x}_t \mid \mathbf{z}_{1:t})$ using the Bayes' rule by

$$p(\mathbf{x}_t \mid \mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_t \mid \mathbf{x}_t) p(\mathbf{x}_t \mid \mathbf{z}_{1:t-1})}{p(\mathbf{z}_t \mid \mathbf{z}_{1:t-1})} \quad (2)$$

where $p(\mathbf{z}_t \mid \mathbf{z}_{1:t-1})$ is the normalization constant, depending on the likelihood function $p(\mathbf{z}_t \mid \mathbf{x}_t)$. For particle filter, the posterior pdf is approximated by a random measure, $\{\mathbf{x}_{1:t}^{(i)}, w_t^{(i)}\}_{i=1}^{N}$, where $\{\mathbf{x}_{1:t}^{(i)}, i = 1, \ldots k\}$ is a set of particles with the associated important weights $\{w_t^{(i)}, i = 1, \ldots N\}$. The samples $\mathbf{x}_t^{(i)}$ are drawn from an importance density $q(\mathbf{x}_t^i \mid \mathbf{x}_{1:t-1}^i, \mathbf{z}_{1:t})$ and weights of the samples are update by

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{p(\mathbf{z}_t \mid \mathbf{x}_t^{(i)}) \mid p(\mathbf{x}_t^{(i)} \mid \mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t^{(i)} \mid \mathbf{x}_{1:t-1}^{(i)}, \mathbf{z}_{1:t})} \quad (3)$$

where $q(\mathbf{x}_t^{(i)} \mid \mathbf{x}_{1:t-1}^{(i)}, \mathbf{z}_{1:t})$ can be chose to be $p(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(i)})$.

In visual tracking, both object and camera motions should be considered. A motion model including the control vector to compensate the global motion (camera motion) can improve the tracking accuracy [12]. At prediction stage, the state vector $\mathbf{x}_t^i$ of the ith particle at time t is predicted by [12]

$$\mathbf{x}_t^{(i)} = \begin{bmatrix} S_{x,t}^{(i)} \\ S_{y,t}^{(i)} \\ W_{x,t}^{(i)} \\ W_{y,t}^{(i)} \\ H_{x,t}^{(i)} \\ H_{y,t}^{(i)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} S_{x,t-1}^{(i)} \\ S_{y,t-1}^{(i)} \\ W_{x,t-1}^{(i)} \\ W_{y,t-1}^{(i)} \\ H_{x,t-1}^{(i)} \\ H_{y,t-1}^{(i)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} G_{x,t} \\ G_{y,t} \end{bmatrix} + \varepsilon_t^{(i)} \quad (4)$$

where $[S_{x,t}^{(i)}, S_{y,t}^{(i)}]^T$ and $[H_{x,t-1}^{(i)}, H_{y,t-1}^{(i)}]^T$ are the position and scale of the target, respectively, $\varepsilon_t^{(i)}$ is Gaussian noise, $[G_{x,t}, G_{y,t}]^T$ is camera motion, and $[W_{x,t}^{(i)}, W_{y,t}^{(i)}]^T$ is the object motion on the 2-D image after compensating the camera motion.

At the correction stage, the object state is corrected using [14], where the weights of particles are computed using color information. Then the current state $\mathbf{x}_t$, is estimated using the minimum mean square error over the posterior

$$E[\mathbf{x}_t] \approx \int \mathbf{x}_t \sum_{i=1}^{N} w_t^{(i)} \delta(\mathbf{x}_t - \mathbf{x}_t^{(i)}) = \sum_{i=1}^{N} w_k^{(i)} \mathbf{x}_k^{(i)} \, . \quad (5)$$

3. DERIVING INTRINSIC IMAGES FROM MIXED IMAGE SEQUENCES

A mixed image, a mid-level description of scenes, can be decomposed into a reflectance image and an illumination image. Based on the sparse prior over derivative filters on the previous T frames, this paper adopts the method in [7] to estimate an illumination image of the current frame and to improve the tracking accuracy on mixed images.

Assume that the reflection is constant while the illumination changes. The method in [7] works in the log domain to recover $l(x, y, t)$ and $r(x, y)$

$$i(x, y, t) = l(x, y, t) + r(x, y) \, , \quad (6)$$

where $i(x, y, t)$ is the mixed image at time t, $l(x, y, t)$ is the illumination image at time t, and $r(x, y)$ is the reflectance image. Given N filters $\{f_n\}$, the output of the filter is

$$o_n(x, y, t) = i(x, y, t) * \{f_n\} \, , \quad (7)$$

and, the filtered reflectance image is denoted by

$$r_n(x, y) = r(x, y) * \{f_n\} \, . \quad (8)$$

With the assumption that the filtered $l(x, y, t)$ of natural images is Laplacian distributed and independent over space and time, the likelihood is defined as

$$P(o_n \mid r_n) = \frac{1}{Z} \exp(-\beta \sum_{x,y,t} | o_n(x, y, t) - r_n(x, y) |) \quad (9)$$

The ML estimate of the reflectance image is

$$\hat{r}_n(x, y) = \text{median}_t o_t(x, y, t) \, . \quad (10)$$

Since

$$f_n * \hat{r} = \hat{r}_n \, , \quad (11)$$

the reflectance image can be obtained using the pseudo-inverse solution

$$\hat{r} = g * (\sum_n f_n^r * \hat{r}_n) \, . \quad (12)$$

where $f_n(x, y) = f_n^r(-x, -y)$ and g is the pseudo-inverse solution

$$g * (\sum_n f_n^r * f_n) = \delta \, . \quad (13)$$

Finally, the illumination image is estimated by

$$\hat{l}(x, y, t) = i(x, y, t) - \hat{r}(x, y) \, . \quad (14)$$

4. THE PROPOSED VISUAL TRACKING SCHEME FOR MIXED IMAGES

To improve the tracking accuracy on mixed images, this section proposes the Weiss mask generated from the intrinsic illumination image to improve the correction stage. The proposed scheme is stated as follows.

1. Selection: Select N random samples from the previous state.
2. Prediction with motion compensation model [12]: Estimate the camera motion using SURF. For each particle, predict its current state using the motion matrix and the control term, i.e. camera motion.
3. Correction: Based on the RGB color histograms of the mixed image and the [I、R-G、Y-B] color histograms, determined by the Weiss mask of the illumination image, optimize the weight of each particle using ML.

The Weiss mask aims at indicating non-reflectance regions on each mixed image. To construct the Weiss mask, the intrinsic illumination image of the current frame is derived by the method in [7], using the previous five frames. In fact, the estimated illumination image includes not

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

Figure 1. Test results of the proposed scheme on video #1. (a) Illumination image. (b) Reflectance image. (c) Weiss mask. (d) Mixed image.

only the transmitted scene (Fig. 1(a)) but part of the reflectance image (Fig. 1(b)), caused by the camera motion or varying lighting. Thus, thresholding and morphology, i.e. erosion, dilation, and erosion, are applied to the illumination image to remove the noisy parts. In Figs. 1(c) and 1(d), the estimated object state is shown on the Weiss mask and the mixed image, respectively.

Regarding color perception, physicists generally agree with the trichromatic theory, the Young-Helmholtz theory [15] of color vision. However, psychologists often agree with opponent-process theory, proposed by Hering [16]. Thus, at the correction stage of the proposed scheme, in addition to RGB color distribution, computed using the method in [17], the [I、R-G、Y-B] color distributions corresponding to the non-reflectance regions on the mixed image, indicated by the Weiss mask, and the rectangular region centered at a particle, are calculated to improve the tracking accuracy. The similarity between the region centered at each particle and the target template are then calculated using the Bhattacharyya distance/coefficient [17])

$$d=\sqrt{1-\sum_{u=1}^{m}\sqrt{p^{(u)}q^{(u)}}} \tag{15}$$

where q is the distribution of the target template, p is the distribution of the candidate, and u is the index of a histogram bin. Assume that the Bhattacharyya distance of the RGB color space and that of the [I、R-G、Y-B] color space are independent, this section defines the likelihood of the nth particle as

$$\pi_t^{(i)}(p,q)=p(\mathbf{z}_t^c \mid \mathbf{x}_t=\mathbf{s}_t^{(i)})\,p(\mathbf{z}_t^W \mid \mathbf{x}_t=\mathbf{s}_t^{(i)}) \tag{16}$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\left(\frac{(dc_t^{(i)})^2}{2\sigma_c^{\,2}}+\frac{(dw_t^{(i)})^2}{2\sigma_W^{\,2}}\right)\right), \tag{17}$$

where $\mathbf{z}_t^c, \mathbf{z}_t^W$ are the measurement state of RGB and [I、R-G、Y-B] color space, respectively, $dc_t^{(i)}$ and $dw_t^{(i)}$ denote the Bhattacharyya distances of the RGB and [I、R-G、Y-B] color histograms, respectively. σ_c , σ_W are the variances of $dc_t^{(i)}$ and $dw_t^{(i)}$, respectively. By taking the log of (17),

$$\begin{aligned}\ln \pi_t^{(i)}(p,q) &= \ln\frac{1}{\sqrt{2\pi}\sigma}-\left(\lambda_t^{(i)}\frac{(dc_t^{(i)})^2}{2\sigma_c^{\,2}}+(1-\lambda_t^{(i)})\frac{(dw_t^{(i)})^2}{2\sigma_W^{\,2}}\right)\\ &= \lambda_t^{(i)}\left(\frac{(dw_t^{(i)})^2}{2\sigma_W^{\,2}}-\frac{(dc_t^{(i)})^2}{2\sigma_c^{\,2}}\right)+const\end{aligned} \tag{18}$$

Finally, given that $\lambda_t^{(i)}$ ranges from 0 to 1, (17) is maximized by

$$\lambda_t^{(i)}=\begin{cases}1, & (dw_t^{(i)})^2/2\sigma_w^{\,2} > (dc_t^{(i)})^2/2\sigma_c^{\,2}\\ 0, & (dw_t^{(i)})^2/2\sigma_w^{\,2} \le (dc_t^{(i)})^2/2\sigma_c^{\,2}\end{cases}. \tag{19}$$

5. EXPERIMENTAL RESULTS

The accuracy of the proposed scheme is evaluated using eight sequences. Samples of videos #1 and #3 are shown on Figs. 2 and 3, and the remains are given in Table I. The eight sequences are including seven mixed sequences and one without reflection. The first sequence, i.e. video #1, a mixed one, is from [5], and the last sequence, i.e. video #8, is from dataset 5 of 2001 PETS [19], without reflection. The other sequences, recorded by NCU VCLab in Taiwan, are captured under different conditions, including varying illumination, static or camera motion, single or multiple objects, and degree of reflection. From videos #1 to #3, the regions of targets on the first few frames do not contain the layer of reflection, whereas such regions on the latter frames always include the layer of reflection. From videos #4 to #7, the regions of targets on the first few frames contain the layer of reflection, whereas such regions on the latter frames may either include the layer of reflection or not. Videos #2 to #7 are downsampled by the factor of 8 in both width and height before tracking. Regarding the test conditions of the proposed scheme, the number of particles is 50. The initial state is detected manually in the mixed image. The number of bins is eight in each color histogram. Variances $\sigma_c^{\,2}$ of and $\sigma_w^{\,2}$ are assumed to be the same.

Few previous tracking methods focus on the reflection problem. Thus, the proposed scheme is compared with the fast L1 tracker [18], extended from [3], a state-of-the-art tracker. For frames with target occlusion, e.g. from t=25 to t=31 in video #1 (Fig. 2), both the proposed scheme and the L1 tracker work well. However, the proposed scheme outperforms the L1 tracker for frames with reflection, e.g. after t=35 in video #1 (Fig. 2). The target region, including the layer of reflection, will be quite different from the target template, constructed from the first frame without the layer of reflection. This will cause the significant difficulty in tracking. For example, for sequences with strong reflection, e.g. video #3, the L1 tracker fails since it cannot tell apart from the reflection and the target (Fig. 3). However, the proposed scheme keeps tracking the target well since it also refers to the reflection information from the Weiss mask. In contrast, if the target template is with reflection whereas the target region is not, it will be much harder to track accurately, e.g. video # 6 and video #7.

To evaluate the tracking accuracy, the comparison of the tracking error with Root Mean Square Error (RMSE) between the proposed scheme (solid line) and the L1 tracker (dotted line) is given in Fig. 4. For each sequence, the value of RMSE at each time instant is the average of ten randomized tests. For the first few frames of all sequences,

Table 1. The first frames of test videos #2 and videos #4-#8.

Video #2	Video #4	Video #5	Video #6	Video #7	Video#8

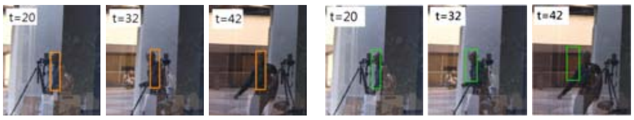

Figure 2. Tracking results for video #1. (a) Proposed scheme. (b) L1 tracker.

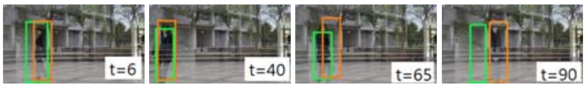

Figure 3. Comparison of the estimated states of the target by proposed scheme (orange) and L1 tracker (green) on video #3.

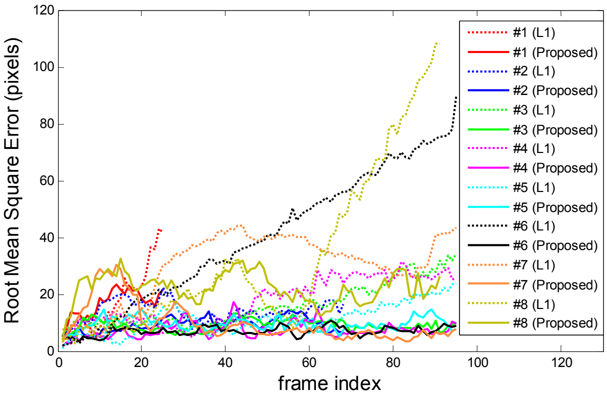

Figure 4. Comparison of the tracking accuracy between the proposed scheme and the L1 tracker [18].

L1 tracker always has smaller RMSE than the proposed scheme. However, the L1 tracker fails in case of reflection while the proposed scheme works well. At the end of all sequences, the proposed scheme always leads to smaller RMSE than the L1 tracker. For video #8 that has no reflection, the proposed scheme is comparable to the L1 tracker. To demonstrate the necessity of consideration of reflection information, the comparison of the tracking error (RMSE) between the proposed scheme (solid line) and the method in [12] (dotted line) is given in Fig. 5. For sequences captured by cameras on mobile platforms, e.g. video #8, the method with the motion compensated model in [12] can improve the tracking accuracy. For sequences with both the strong reflection and camera motion, e.g. video #7, the proposed scheme performs well since it significantly improves the method in [12] by revising the correction stage.

Finally, our analysis finds that most of the covariances of [18], [12], and the proposed method on videos #1-#8 approach zero. Thus, the variances of the estimated x and y positions of the target on the last frame of each video are

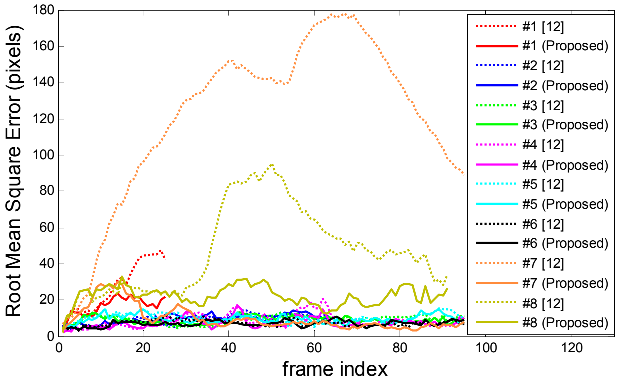

Figure 5. Comparison of the tracking accuracy (RMSE) between the proposed scheme and the method in [12].

Table 2 (σ_x^2, σ_y^2), the variances of the estimated x and y positions of the target on the last frame of each video.

	Video #1	**Video #2**	**Video #3**	**Video #4**
[18]	(133,24)	(4,1)	(94,1)	(19,1)
[12]	(475,95)	(7,3)	(4,10)	(6,11)
Proposed	(135,108)	(5,9)	(2,10)	(7,9)
	Video #5	**Video #6**	**Video #7**	**Video #8**
[18]	(60,23)	(1,1)	(466,31)	(9,4)
[12]	(23,38)	(2,3)	(85,56)	(519,297)
Proposed	(5,24)	(1,2)	(15,4)	(20,42)

given in Table II. Compared with the method in [18] leading to small tracking errors on videos #2-#5 (Fig. 4), the estimated positions of the proposed method is more stable. Compared with the method in [12] that leads to small tracking errors on videos #2-#6 (Fig. 4), the estimated positions of the proposed method is also more stable.

6. CONCLUSIONS

Different from previous visual tracking methods, this paper proposes a novel method to improve the visual tracking accuracy in case of specular reflections in real world. The correction stage of PF is enhanced by maximizing the likelihood for each particle using both the color information of mixed images and intrinsic images. Experimental results show that the proposed scheme performs well in case of occlusion and reflection. In future work, information fusion at the correction stage of particle filter will be investigated to improve the reliability and robustness of the proposed scheme.

7. ACKNOWLEDGEMENT

This work is supported by National Science Council of Taiwan under the grants NSC-101-2221-E-008-060. In addition, we deeply appreciate the discussions of the blind source separation with Mr. Shao-Yu Tan, a graduate student with Institute of Communications Engineering at National Chiao Tung University, Hsinchu, Taiwan.

8. REFERENCES

[1] Y. Wu and T. Huang, "Robust Visual Tracking *by Integrating Multiple Cues based on Co-inference Learning," Int'l Journal of Computer Vision*, vol. 58, no. 1, pp.55–71, Jun. 2004.

[2] Y. Jin and M. Farzin, "Variational Particle Filter for Multi-object Tracking," *in Proc. IEEE International Conference on Computer Vision*, pp.1-8, Rio de Janeiro, Brazil, Oct. 2007.

[3] X. Mei and H. Ling, "Robust Visual Tracking Using L1 Minimization," *in Proc. IEEE International Conference on Computer Vision*, , pp.1436-1443, Barcelona, Spain, Sept. 2009.

[4] H. Farid and E. Adelson, "Separating Reflection and Lighting Using Independent Component Analysis," *in Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 262-267, Fort Collins, USA, June 1999.

[5] B. Sarel and M. Irani, "Separating Transparent Layers through Layer Information Exchange," *in Proc. European Conf. Computer Vision*, pp.328-341, Prague Czech Republic, May 2004.

[6] K. Hara, I. Kohei, and U. Kiichi, "Separation of Layers from Images containing multiple reflections and transparency using cyclic permutation," *in Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, pp.1157-1160, Taipei, Taiwan, Apr. 2009.

[7] Y. Weiss, "Deriving Intrinsic Images from Image Sequences," *in Proc. IEEE International Conference on Computer Vision*, vol. 2, pp.68–75, British Columbia, Canada, July 2001.

[8] A. Levin and Y. Weiss, "User Assisted Separation of Reflections from A Single Image Using A Sparsity Prior," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1647-1655, Sept. 2007.

[9] A. Levin, A. Zomet, and Y. Weiss. "Separating Reflections from A Single Image Using Local Features," *in Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I-306-I-313, Washington, USA, June 2004.

[10] K. Gai, Z. W. Shi, and C. S. Zhang, "Blind Separation of Superimposed Moving Images Using Image Statistics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 19-32, Jan. 2012.

[11] M. A. Elgharib, F. Pitie, A. Kokaram, and V. Saligrama, "User-Assisted Reflection Detection and Feature Point Tracking," *in Proc. European Conference on Visual Media Production*, Nov. 2013.

[12] J.-Y. Lu, Y.-C. Wei, and C.-W. Tang, "Visual Tracking Using Compensated Motion Model for Mobile Cameras," *in Proc. IEEE International Conference on Image Processing*, pp. 489-492, Brussels Belgium, Sept. 2011.

[13] N. Gordon, M. Arulamalam, S. Maskell, and T. Clapp, "A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking," *IEEE Trans. Signal Processing*, vol. 50, no.2, pp. 174-188, Feb. 2002.

[14] J. L. Yang, D. Schonfeld, and M. Mohamed, "Robust Video Stabilization Based on Particle Filter Tracking of Projected Camera Motion," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 19, no. 7, pp. 945-954, July 2009.

[15] T. Young, The Bakerian lecture: On The Theory of Light And Colours, *Phil. Trans. R. Soc. Lond. 92*, pp. 12–48, 1802. (doi:10.1098/rstl.1802.0004)

[16] D. H. Hubel, Eye, Brain, and Vision, Scientific American Library Series (Book 22), W. H. Freeman, 2nd Edition, New York, USA, May 15, 1995.

[17] K. Nummiaro, E. Koller-Meier, and L. V. Gool, "An Adaptive Color Based Particle Filter," *Image and Vision Computing*, vol. 21, no. 1, pp.99-110, January 2003.

[18] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real Rime Robust L1 Tracker Using Accelerated Proximal Gradient Approach," *in Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1830-1837, Providence Rhode Island, June 2012.

[19] PETS'2001 The Second IEEE International Workshop on Performance Evaluation of Tracking and Surveillance December 9, 2001 (in conjunction with IEEE CVPR'2001) Kauai, Hawaii, USA.