Cast Shadow Detection Based on Semi-supervised Learning

Salma Kammoun Jarraya[1], Rania Rebai Boukhriss[1], Mohamed Hammami[2], and Hanene Ben-Abdallah[1]

[1] MIRACL-FSEG, Sfax University, Rte Aeroport Km 4, 3018 Sfax, Tunisia
[2] MIRACL-FS, Sfax University, Rte Soukra Km 3 BP 802, 3018 Sfax, Tunisia
{Salma.Kammoun,Hanene.Benabdallah}@fsegs.rnu.tn,
Rania.Rebai@hotmail.fr, Mohamed.Hammami@fss.rnu.tn

Abstract. In this paper, we tackle the shadow problem in depth for a better foreground segmentation. We propose a novel variant of co-training technique for shadow detection and removal in uncontrolled scenes. This variant works according to a powerful temporal behavior. Setting co-training parameters is based on an extensive experimental study. The proposed co-training variant runs periodically to obtain more generic classifier, thus improving speed and classification accuracy. An experimental study by quantitative, qualitative and comparative evaluations shows that the proposed method can detect shadow robustly and remove the 'cast' part accurately from videos recorded by a static camera and under several constraints.

Keywords: Cast shadow detection and removal, foreground segmentation, semi-supervised learning, co-training technique.

1 Introduction

Foreground segmentation is a cornerstone step in computer vision applications. On the other hand, every efficient foreground segmentation method must include special processing to overcome the difficulties behind moving shadow detection and removal. Moving shadows occur by partial or entire occlusion of direct light from a light source by a moving object. In natural scenes, moving shadows are attached to foreground objects, which make them un-distinguishable by foreground segmentation methods. This in turn reduces the efficiency of the foreground segmentation methods.

The complexity of shadow detection is due to the dynamic changes in natural scenes (single/multiple or natural/artificial light sources, illumination changes, camera position, scene geometry information, etc) and to the visual shadow features (color darken to clear, shape, size). To apprehend this complexity, the proposed shadow detection methods use either a statistical approach or a deterministic approach to classify a pixel as foreground or shadow. In the statistical methods, pixel classification is based on probabilistic functions; in addition, they can be parametric (*cf.* [1][2]) or non-parametric (*cf.* [3][4]). On the other hand, the determinist methods classify pixels

A. Campilho and M. Kamel (Eds.): ICIAR 2012, Part I, LNCS 7324, pp. 19–26, 2012.

through an on/off decision in two ways: 1) based on a model [5][6] which require a priori knowledge of the objects and/or the scene; 2) based on empirical thresholds [7][8] (without using a model). The main drawback of both statistical Parametric and Determinist Model based methods is the difficulty of defining the parameters/models. Deterministic Non-Model based methods are simples and fasts but lack generality. This difficulty is overcome by Statistical Non-Parametric (SNP) methods by statistically selecting the classification thresholds.

In this paper, we propose a novel SNP method based on a **S**emi-**S**upervised **L**earning (**SSL**) scenario. Our method on-line and periodically generates a prediction model used for shadow detection in uncontrolled scenes. It presents two main contributions: The first contribution is a new variant of co-training technique for SSL, which: 1) reduces the run-time by running periodically our co-training process according to a novel temporal framework; and (2) generates a more generic prediction model for a more accurate classification. The second contribution is an optimal co-training setting given by an extensive experimental study.

The remainder of this paper is structured as follows. In section 2, we present our proposed method for moving shadow detect and removal. The proposed co-training variant is detailed in section 3. The efficiency and accuracy of our work are illustrated by an exhaustive experimental evaluation and comparison results in Section 4. Finally, a summary and ongoing works are presented in Section 5.

2 Proposed Method

Our method aims at detecting shadow pixels from moving pixels and eliminating the cast part. It operates in three major steps: **(1)** Moving pixel classification into shadow/non-shadow to obtain shadow pixels part; **(2)** Shadow Pixel Classification into cast/self shadow; and **(3)** Building moving Pixels Card by grouping non-shadow pixels and self-shadow pixels. The input of **Moving Pixel Classification** is moving pixels. These latter can be identified through a fast and accurate moving pixel detection method based on background modeling approach described in our paper [9]. This method is demonstrated robust and accurate under most of the common problems in foreground segmentation. The goal of this first step is to discriminate foreground pixels from shadow pixels. For this, it classifies each candidate moving pixel according to a prediction model (*PM*) into shadow/non-shadow pixel to build two masks M_s and M_{ns} that denote respectively masks of shadow/non-shadow pixels. The non-shadow pixels (M_{ns}) correspond to the real moving pixels. The prediction model generation framework is described in section 3.

The basic idea of **Shadow Pixel Classification** is to classify each shadow pixel ($M_s(x,y)=1$) into self/non-self by *PM_Self*. *PM_Self* is obtained in our previous work [10]; this model shows effectiveness and robustness to deal with self/cast shadow pixel classification. Self and non-self shadow pixels assigned respectively *1* and *0* to obtain two masks, non-self shadow mask that correspond to cast shadow mask (M_{cast}) and self shadow mask $\left(M_{self}\right)$.

A Moving Pixels Card (*MPC*) is obtained in the third step **"Building moving Pixels Card"** by applying a logical OR between the non-shadow mask (M_{ns}) and the self shadow mask $\left(M_{self}\right)$.

3 Prediction Model Generation

Prediction model generation constitutes a considerable work due to the strong similarity between visual features of foreground pixels (non-shadow pixels) and shadow pixels. We propose to generate prediction model (*PM*) based on **S**emi-**S**upervised **L**earning (**SSL**). SSL is a combination of supervised and unsupervised learning where typically a small labeled set and a large unlabeled set are used for training. According to conclusions of previous works [11][12], we can summarize the great motivation, in literature, to use SSL in two reasons: firstly labeling of a huge set of instances by skilled human expert can be a time-consuming task and, secondly, it has been shown that using unlabeled data for learning improves the accuracy of the produced classifier. We are interested to the second motivation, since visual features of data to be classified (moving pixels) are tightly influenced by changes in natural scene, thus classification task requires often adaptation of the classifier. Evidently, adaptability is assured, in SSL, by the unlabeled data using in learning step.

There are different algorithms for semi-supervised learning among which we find the co-training technique. In fact, we can sum up advantages of co-training learning in the following three benefits: (i) gives better accuracy, (ii) needs fewer labeled data, and (iii) requires less training time. In classical variant of co-training algorithm, two classifiers are trained using two views of the small labeled data. Then each of them assigns labels to all unlabelled examples, selects the most confidently predicted and adds them to the labeled pool, and, after N rounds, when training is completed, the label of new instances is predicted by the classifier that is more confident on the example. Even classical co-training variant has been successfully applied to a number of classification fields, such variant cannot deal robustly with shadow/non-shadow classification task. In fact, if we consider a video flux, from one frame it is possible to generate classifiers based on the enlarged labeled pool after N iterations. So this pool cannot include huge significant labeled moving pixels since in natural video flux, foreground and shadow pixels features might change frequently with the dynamic scene conditions. Thus, we propose a novel variant of co-training technique relies on significant labeled pool and extract more generic useful knowledge. This variant works according to a powerful temporal behavior (Fig.1). In fact, let *P* and Q denote partials periods of the flux. Two weak classifiers are re-trained N rounds on each frame of a pre-defined period (*P frames*) to **(i)** build a huge significant labeled pool (**MGLP**: **M**odel **G**eneration **L**abeled **P**ool). Based on this pool, **(ii)** prediction model (PM_i) is obtained by a learning technique. Between frames $P+1$ and $Q+P$, labels of moving pixels are predicted by (PM_i). For period $[Q+1,Q+P]$, both classifiers are invoked to update back MGLP by samples related to the actual conditions and in frame $Q+P$, another PM_i is generated. System continues to run with the same rules. This temporal behavior of co-training makes the classifier PM_i more generic allowing stopping co-training process after a few periods. Next, we details steps for prediction model generation.

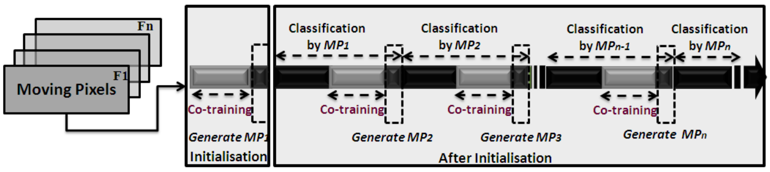

Fig. 1. Temporal behavior of our method

3.1 Building MGLP

The general principle of building MGLP is presented by Fig.2. For the initialization phase, two classifiers (*SVM1* and *SVM2*) are trained relies on two different feature sets of small human-labelled data set (**LLP**: **L**earning **L**abelled **P**ool) and a large unlabelled data set (input moving pixels). For the periods following the initialization phase (when size of MGLP>*M*), the initial *M* pixels of the LLP is randomly selected from MGLP. The two classifiers are re-trained *N* rounds on each frame of a pre-defined period (*P frames*). After a training round, the most confident labelled examples from each class and their 8-neighboorhood are added to: LLP and MGLP. Unlike classical co-training variant, we increase the certainty of the most confident labelled examples selection by involving classification decision of a prediction model (***PM_Shadow***) and applying a second prediction model (***PM_Self***) to the most confident shadow pixels (shadow class) to discriminate cast shadow pixels and to distinguish them from self shadow pixels. In fact, *PM_Shadow* is obtained by an off-line supervised learning. For these, firstly, we build a training set (*1897998* shadow pixels and *1791562* non-shadow pixels) from famous indoor and outdoor sequences recorded in typical conditions. Secondly, we identify the effectiveness shadow pixels features in order to build n-dimensional table from our training corpus. In fact, reading through the literature the most features[1] exploitable for describing shadow pixels are: *Brightness distortion, Normalized Cross Correlation*, *Ratio(Im,BG)* in *V* (HSV color space), *Ratio(Im,BG) in R*, *G* and *B* (RGB color space), *Difference(Im,BG)* in *H* and *S* (HSV color space), *Difference(Im,BG)* in *Y* (YCrCb color space), and *Edge magnitude* and *gradient distortion*. The eleven shadow features are considered in our work. Thirdly, we choose the appropriate supervised learning techniques to generate ***PM_Shadow***. In literature, there are several techniques of supervised learning, each having its advantages and drawbacks. So, among the most important criteria to compare supervised learning techniques is the comprehensibility of the learned model which leads us to a well-accepted technique, that is, the induction of decision trees [13]. Six data mining algorithms were studied according to **C**lassification **A**ccuracy (***CA***) and **F**alse **C**lassification **R**ates (***FCR*** complement of *CA*), including *ID3, CR.5, Cost-Sensitive C4.5, One-Vs-All Decision Arbre, A limited search* and *Improved Chaid*. It seems that the best *CA* (9.34%) and *FCR* (0.66%) rates were achieved by C4.5.

---

[1] *Im* and *BG* denotes pixel values respectively in Frame and Background.

Although the general principle of building MGLP is one consideration when choosing to use co-training, we also look for setting the parameters of our system to improve running time and increase accuracy. According to our several experimental studies, even we start by a small training set (here *M=100* pixels in LLP), classification results may be improved by automatically labeled examples. Secondly, number of rounds is one important consideration for execution time, as the number of rounds is small (here *N=5*) as we improve speed. This does not means that the system decreases in classification accuracy. Co-training process requires that features set must be splitting according to conditionally independent assumption. However, this assumption is violated in major real field. And even if the assumption holds, most of the time there are few possible mutual information's between features. Obviously, we can split features used to generate *PM_Shadow* into nine illumination based features and two edge based features. Evidently, Edge magnitude and gradient distortion constitute the first features set (*<f1,f2>*), therefore, illumination based features is the second set (*<f'1,f'2,…>*). Nevertheless, because the separation of the training data by SVM is often easier achieved in a higher dimensional space and due to computational expensive to examine all features, we are selected by *ReliefF* algorithm pertinent illumination based features. In fact, there are many different algorithms for features selection; we are interested to *ReliefF* algorithm since it is unaffected by feature interaction [14]. *Difference(Im,BG)* in *H* and *S* (HSV color space) are the two best ranked features, thus they constitute the second features set (*<f'1,f'2>*).

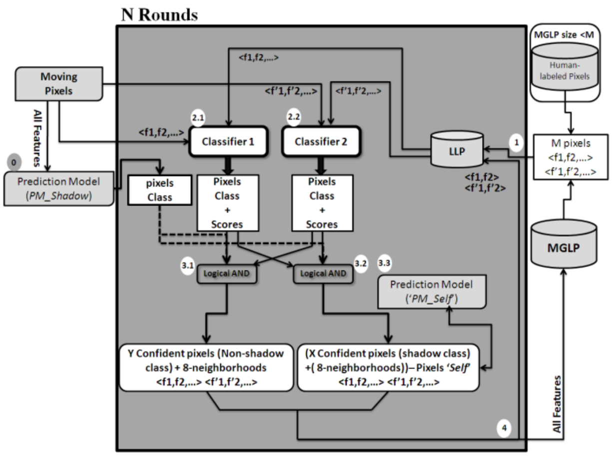

Fig. 2. Process to Build MGLP

3.2 Generating Prediction Model

In classical co-training algorithm, the classifier that is more confident on the example to be labeled is invoked for classification [15]. For several raisons, in our work, the classifier used to generate on-line a prediction model is C4.5 algorithm [16]. Firstly, according to our experimental study in supervised learning, C4.5 gives best classification accuracy in case of largest pool (MGLP). Secondly, even with small pool, speed accuracy with SVM is better, so in case of largest pool C4.5 is more adopted for better speed accuracy. Obviously, general rule of SVM classification make it scarcely find a separating hyperplane for complex data in largest pool. Thirdly, decision trees generate rules. A rule is a conditional statement that can easily and rapidly use for classification.

4 Experimental Results

In order to evaluate our proposed method, we carried out a series of experiments. Firstly, we present our quantitative and qualitative results. Secondly, we have compared our contribution with three well-known methods of shadow detection and removal given in a comparative reviewing [17]: (1) SP based method [18]; (2) SNP based method [3] and (3) DNM based method [8]. Each method is considered as reference in its category. A complementary comparative study is performed with two recent works in statistical approach: (4) based on parametric models [19] and (5) SNP based on SSL process [4]. We performed experiments on a dataset that contains a set of famous indoor and outdoor sequences recorded in typical conditions. (i) *HighwayI* and *IntellegentRoom* are used to evaluate performance of the moving pixel classification by quantitative scores. (ii) *HighwayIII* and *Compus* mainly used for qualitative shadow detection since typical Ground-Truth frames for these sequences are not available. For the performance evaluation, we have used the mainly used metrics (Detection accuracy (η) and Shadow discrimination accuracy (ξ)) to judge the effectiveness of the shadow detection methods, these metrics was proposed and detailed in a well-know survey paper [17] in this field. In addition to the above quantitative metrics, we also consider in our evaluation the Classification accuracy (*CA*) and Precision of moving pixel detection. Figure 3(A) presents respectively for '*HighwayI*' and '*Intelligent Room*' sequences precisions rates of moving pixel detection before and after cast shadow detection and removal. Curves show clearly improvements of ≈ 0.27 percent recorded by our method for moving pixel detection (see also Fig. 3(B)). Qualitative results are given by Fig. 4 for significant frames from (A)'*HighwayIII*' and (B)'*Compus*'. These results show the adaptive behavior robustness of our method. In fact, our method works independently of the dynamic environment and variety of scenes.

These latter results encouraged us to further experiment by comparing our contribution to best results of each shadow detection approach represented by well-known

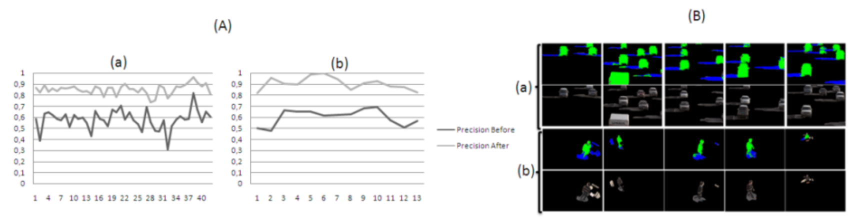

Fig. 3. Obtained (A) moving pixel precision rates Before and After shadow detection and removal and (B) Qualitative results for (a) *HighwayI* sequence and (b) *IntellegentRoom* sequence.

methods ([18], [3] and [8]). As we can see in Fig.5(A), for '*Intelligent Room*' sequence, even our method echoed similar (ξ) rates (≈89%) with best ones (90%) given by methods (1) and (3), we record the best (η) rate of 84.85 percent . While for *HighwayI*' sequence, our method echoed the best rates of both metrics (ξ=93.47% and η= 86.32%). Comparative study with recent works ([19] and [4]) (Fig. 5(B)) shows that our method gives, for the two sequences, the best rates of (ξ) and (η) compared to methods (4) and (5). The two latter methods record (ξ) rates between 66.54 percent and 89.76 percent and (η) rates between 60.24 percent and 83.44 percent. Indeed, since the method (5) is based on SSL with a classical algorithm of co-training executed in each frame for the detection and classification. The gain is not only for (η) and (ξ) rates but also for the speed accuracy due to, the temporal behavior and the setting of our co-training variant.

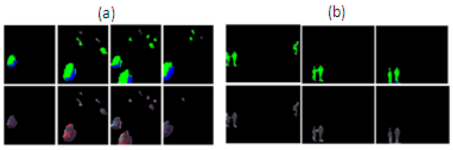

Fig. 4. Qualitative results (a) for *HighwayIII* and (b) for *Compus*

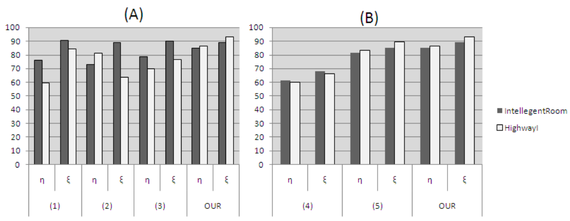

Fig. 5. Comparative results with (A) well-knows methods: (1) SP [18], (2) SNP [3] and (3) DNM [8] and (B) Recent statistical methods: (4) SP [19] and (5) SNP [4]

5 Conclusion

In this paper, we presented method for shadow detection and removal based on a semi-supervised learning by a new co-training variant. The proposed co-training variant was evaluated by a series of experiments with various sequences against different conditions. We obtain a good compromise between shadow detection accuracy and shadow discrimination accuracy rates with percentages between 84.85 percent and 93.47 percent. Future works will focus on evaluate our method with upper computer vision applications steps.

References

1. Wang, Y., Wang, S.: Shadow Detection of urban color aerial images based on partial differential equations. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVII, part B2 (2008)
2. Dian, W.: The research on a mixtrue gausssian based clustering algorithm of make background model and supressing shadow. Northwestern Polytechnical University (2006)

3. Horprasert, T., Harwood, D., Davis, L.S.: A statistical approach for real-time robust background subtraction and shadow detection. In: ICCV Frame-Rate Workshop, vol. 99, pp. 1–19 (1999)
4. Joshi, A.J., Papanikolopoulos, N.P.: Learning to Detect Moving Shadows in Dynamic Environments. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(11), 2055–2063 (2008)
5. Finlayson, G.D., Hordley, S.D., Drew, M.S.: Removing Shadows from Images Using Retinex. In: The 10th Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications, Scottsdale, Arizona, pp. 73–79 (2002)
6. Wang, J.M., Chung, Y.C., Chang, C.L.: Shadow detection and removal for traffic images. In: IEEE International Conference on Networking, Sensing and Control, vol. 1, pp. 649–654 (2004)
7. Liu, H., Yang, C., Shu, X., Wang, Q.: A new method of shadow detection based on edge information and HSV color information. In: 2nd Conference on Power Electronics and Intelligent Transportation System, vol. 1, pp. 286–289 (2009)
8. Cucchiara, R., Grana, C., Piccardi, M., Prati, A.: Detecting objects, shadows and ghosts in video streams by exploiting color and motion information. In: Proceedings 11th International Conference on Image Analysis and Processing, pp. 360–365 (2001)
9. Hammami, M., Jarraya, S.K., Ben-Abdallah, H.: On line Background Modeling For Moving Object Segmentation in Dynamic Scenes. Multimedia Tools and Applications Journal (2011); available in SpringerLink, appear first on-line
10. Jarraya, K.S., Ghorbel, A., Chaouachi, A., Hammami, M.: ROADGUARD Highway Control and Management System. In: Proceedings of the International Conference on Computer Vision Theory and Applications, pp. 632–637 (2011)
11. Schohn, G., Cohn, D.: Less is More: Active Learning with Support Vector Machines. In: Proceedings of the Seventeenth International Conference on Machine Learning, pp. 839–846 (2000)
12. Guggenberger, A.: Semi-supervised Learning With Support Vector Machines. Technischen Universität Wien (2008)
13. Quinlan, J.R.: Induction of Decision Trees. Machine Learning 1(1), 81–106 (1986)
14. Kononenko, I.: Estimating attributes: Analysis and extensions of relief. In: Proceeding ECML 1994 Proceedings of the European Conference on Machine Learning on Machine Learning, pp. 171–182 (1994)
15. Zhu, X.: Semi-Supervised Learning Literature Survey, New York. Technical Report (2008)
16. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kanfmann Publishers, Inc., Los Altos (1993)
17. Prati, A., Mikic, I., Trivedi, M.M., Cucchiara, R.: Detecting moving shadows: algorithms and evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 918–923 (2003)
18. Mikic, I., Cosman, P.C., Kogut, G.T., Trivedi, M.M.: Moving shadow and object detection in traffic scenes. In: 15th International Conference on Pattern Recognition, pp. 321–324 (2000)
19. Joshi, A.J., Atev, S., Masoud, O., Papanikolopoulos, N.: Moving Shadow Detection with Low- and Mid-Level Reasoning. In: IEEE International Conference on Robotics and Automation, pp. 4827–4832 (2007)